TELEVISION

TELEVISION

THE ELECTRONICS OF IMAGE TRANSMISSION

BY

V. K. ZWORYKIN, E.E., Ph.D.
Director of Electronic Research, Research Laboratories
Radio Corporation of America

AND

G. A. MORTON, Ph.D.
Electronics Research Laboratory, RCA
Manufacturing Company, Camden, N. J.

FIFTH PRINTING

NEW YORK
JOHN WILEY & SONS, INC.
LONDON: CHAPMAN & HALL, LIMITED

FIFTH PRINTING, JUNE, 1948

Printed in the U. S. A.

PREFACE

Electronic television, although a relatively new field, has already acquired an extensive literature in the various scientific and engineering journals. As yet, however, little of this material has been presented in book form. The student or engineer desiring a knowledge of the field as a whole is therefore faced with the arduous and somewhat discouraging task of critically examining the many published articles relating to television. It was felt that a detailed survey of this newest of practical means of communication had become almost a necessity. To fill this need, the authors have attempted to select and integrate relevant portions of the available material into a single volume.

The first part of the text is devoted to a consideration of the fundamental physical phenomena involved in television, that is, emission of electrons, fluorescence, electron optics, etc. Part II deals broadly with the field of television as a whole, taking up the relationship between the physical system and the quality of the picture, the principles of ultra-high-frequency transmission and reception of television signals, and the more important methods of pickup and reproduction of images. There follows in Part III an analysis of the components of the electronic television system based upon the storage principle, treating in turn the Iconoscope, Kinescope, electron gun and associated circuits, together with a description of the television transmitter and receiver. The concluding part of the book is devoted to a description of a working television system, as exemplified by the equipment used in the RCA-NBC television project.

For the benefit of students and others having access to laboratories, the description of the preparation of Iconoscopes, Kinescopes, electron guns, etc., has been made sufficiently complete to permit the construction of working models of these devices in an elementary form. Such experience in the experimental techniques involved should be of no little value to those preparing to enter the fields of electronic or television research.

No single volume can hope to cover completely a field as broad as the one in question. The authors in compiling the subject matter were faced with the necessity of leaving out a great deal which is both interesting

and important, and of emphasizing certain phases at the expense of others. Because of this, the story of the growth of television has been omitted, reference being made to earlier developments only when they help to clarify present problems. Furthermore, greater stress has been laid upon the electronic television systems employing the storage principle than upon non-storage systems.

Only through the generous cooperation of the engineering staff of the RCA organization has it been possible to assemble the material found in these pages. In acknowledging this indebtedness, the authors wish to point out that the information obtained from this source was necessarily interpreted in the light of their own experience. Material drawn from various published work has also been used freely throughout the book. While bibliographies are included, they are necessarily incomplete, and, for the convenience of the reader, summaries and reviews have in some instances been cited, rather than the primary source of the information. Finally, the authors gratefully acknowledge their obligation to Dr. E. G. Ramberg, not only for his many contributions to the mathematical treatment of various phases of the subject, but also for his assistance in editing the manuscript and in preparing the material for the printer.

V. K. Zworykin
G. A. Morton

Camden, N. J.
December, 1939

CONTENTS

PART I. FUNDAMENTAL PHYSICAL PRINCIPLES

PART I

Fundamental Physical Principles

CHAPTER 1

EMISSION OF ELECTRONS FROM SOLIDS

Historically, the first recorded observations on electricity in the form of static charges date back to antiquity. In this respect electricity is in no sense a modern discovery. However, it was not until the eighteenth century that electrical conduction and electric current, which form the basis of most of its present usefulness, became known and recognized. During the nineteenth century rapid strides were made by such men as Maxwell, Faraday, Hertz, and many others towards unraveling the laws of electromagnetic phenomena. The last two decades of that century were particularly fruitful, culminating in the discovery of the electron by J. J. Thomson in 1897.

Before the discovery of the electron itself, many observations had been made on phenomena connected with electron emission.

The discovery of the photoelectric effect, that is, the emission of electrons under the action of light, dates back to observations made by Hertz in 1887 on the fact that electrical discharges were facilitated if the negative electrode of the spark gap was irradiated with ultra-violet light. The following year Hallwachs undertook a systematic study of the effect observed by Hertz and came to the conclusion that negative electricity leaves a body which is illuminated by ultra-violet light. Elster and Geitel carried the work still further and gathered data on the rate at which charge left different metals, and upon the relationship between the wavelength of the incident radiation and the rate of transfer of charge. With the discovery of the electron the nature of the phenomenon was disclosed. Again, in 1905 the theory of photoemission made another great step forward. This occurred when Einstein applied the then new quantum theory, put forward by Planck, to the problem of the photoelectric long-wavelength threshold. Since then the theory has been extended steadily, keeping pace with an increasing knowledge of the quantum mechanics of the solid state. At the same time experimental technique and empirical knowledge have advanced constantly and continue to progress.

Although the effect of temperature on the rate of discharge of electrified bodies had been known to science for a long time, thermionic emission cannot be said to have been discovered until 1883. In this year Edison discovered that a negative current could be made to flow from

an incandescent filament in an evacuated bulb. By the beginning of the twentieth century considerable data had been collected pertaining to this phenomenon. In 1905 O. W. Richardson derived, with the aid of classical thermodynamics, an equation which described the observations fairly accurately. Later the equation derived by Richardson was revised and reinterpreted by Dushman, Langmuir, and others, but the original exponential form was retained. With the development of quantum mechanics the understanding of the effect again took a great stride forward. Nordheim and others, with the aid of the Sommerfeld model of metals, were able to show that thermionic emission is closely related to the phenomena of conduction, photoelectric emission, thermoelectricity, etc. This, together with technical advances in making thermionic cathodes, has brought the subject to its present status.

Secondary emission, that is, the emission of electrons from a surface which is bombarded by high-velocity electrons, did not enter the field until somewhat later. The phenomenon became known about 1900, as a result of experimental work by Starke, Lenard, Hull, O. von Baeyer, and others. Until about 1920 very little was done toward investigating this phenomenon. During the past two decades, however, its practical importance has become more widely recognized, and this has given an impetus to the investigation of this type of emission. At present there is available quite a good deal of reliable empirical data relating to the effect, but as yet no fully satisfactory theoretical interpretation has been forthcoming. However, since the need for this analysis has become very real, there is every reason to believe that the work may be done in the near future. Already a start has been made by H. Fröhlich, who, on the basis of quantum physics, has treated the case of a clean metal surface for limited ranges of bombarding velocities.

In the succeeding sections these phenomena will be discussed without reference to the historical order of their development. For a complete treatment of the history of electron emission, the reader is referred to any textbook on the history of electricity.

1.1. Structure of Metals. Reading across the periodic table of elements from right to left, the elements are found to become increasingly electropositive. That is, the elements in the columns on the left have an electron configuration such that they can easily lose one or more of these electrons and become positive ions. In the solid state these electropositive elements form the group of substances called metals. This group is characterized by the fact that its members are excellent conductors of both heat and electricity. Furthermore the metals are in general both malleable and ductile, although there are some exceptions.

It should be noted that these exceptions occur in the center columns of the table, and are all elements which are only slightly electropositive; in other words, these exceptions might be termed "poor metals."

X-ray analysis of the metals shows them to be crystalline. The crystal structure of these substances (at least for all the good metals) is extremely simple. All the elements of the first column of the periodic table, together with W, Fe, Mo, and many others as well, have a body-centered cubic structure. The very malleable and ductile metals such as Au, Ag, Cu, Al, etc., are face-centered cubic, while others such as Zn, Mg, and Be are hexagonal close-packed.

The atoms making up the metal lattice no longer retain the loosely bound electrons associated with them in their free state. The electrons released when a metal crystal is formed do not have definite positions in the lattice, but exist rather as an electron gas which is free to move through the structure.

This theory, which assumes that the valence electrons form an electron gas, was proposed as early as 1905 by H. A. Lorentz. He assumed that the electrons which form this gas are in thermal equilibrium with the atoms of the lattice, and that they have a Maxwellian energy distribution. On the basis of this hypothesis, in order that the electrons stay within the metal, it is necessary to postulate a potential barrier at the boundaries of the metal. At a large distance from the metal this potential is assumed to be due to the image force of the electron, but since the ideal image force rises to infinity at the surface of the metal, it is necessary to assume a departure from this law near the boundary.

The Lorentz hypothesis explained fairly satisfactorily electrical conduction, thermionic emission, and a number of other observed phenomena, but required that the specific heat of a metal be nearly twice the observed value.

It was not until the advent of quantum mechanics and the exclusion principle that the reason for this discrepancy became evident. With the aid of the new mechanics Sommerfeld, Pauli, and others were able to modify the Lorentz theory in such a way that it gave a very accurate quantitative account of the behavior of metals.

The difference between the old Lorentz theory and the Sommerfeld theory lies in the type of energy distribution assumed for the electron. In either case the energy E of an electron is:

$$E = V_0 + \frac{p^2}{2m}, \tag{1.1}$$

where V_0 is its potential energy and p its momentum. Lorentz placed no restrictions upon the values which p could assume. The newer theory,

on the other hand, assumes, in accordance with quantum laws, that the three momentum components are restricted and, for a block of metal with the dimensions L_x, L_y, and L_z, can only have values satisfying the relation:

$$p_x = \frac{n_x h}{2\pi L_x}, \quad p_y = \frac{n_y h}{2\pi L_y}, \quad p_z = \frac{n_z h}{2\pi L_z}, \tag{1.2}$$

where h is Planck's constant and n_x, n_y, and n_z are integers. The allowed energies are therefore not continuous but consist of a vast number of discrete levels. A further restriction is imposed by the exclusion principle. This rule states that only two electrons can exist in any one energy level, two electrons being allowed because of the two possible spin directions.

The effect of these two restrictions becomes clearly apparent upon a consideration of what happens when the crystal is cooled to absolute zero. In the older theory, all the electrons come to rest, since they are in thermal equilibrium with the atoms making up the crystal lattice. However, this cannot happen according to the newer theory because there can be only two electrons in a state of zero energy. The remaining electrons must retain enough energy to be able to occupy other energy levels. As a result all the lower levels are filled up, so that energies corresponding to several electron volts exist in a crystal even at zero absolute.

If, instead of being at zero absolute, the crystal is at room temperature, the mean energy of its atoms is still far below the energy held by the electrons in the highest filled level. Consequently the probability of an electron receiving enough energy from the thermal motion of the atoms making up the crystal to raise it to an unfilled energy level is extremely small. The electrons, therefore, because they can be given almost no thermal energy by the lattice, contribute very little to the heat capacity of a metal. The greatest difficulty encountered by the older theory is thus overcome.

At very high temperature the mean energy of the atoms making up the lattice may become greater than the energy of the highest filled level. Under these circumstances the electrons receive thermal energy and come into equilibrium with the lattice, their energy distribution approximating the Maxwellian distribution postulated by the earlier theory. An abnormally high specific heat would exist at such temperatures. A calculation of the temperature required to increase the specific heat appreciably shows it to be beyond the vaporizing temperature of any known metal.

A more quantitative consideration, starting with the potential barrier, is necessary if these ideas are to be of practical value. The potential energy, which keeps the electrons within the material, increases with the

distance from the surface, asymptotically approaching some arbitrarily chosen zero at infinity. As has been mentioned, at distances from the crystal greater than the lattice spacing this potential is the result of the simple image force. This force, as is well known, is given by the relation $F = -e^2/4x^2$, where x is the distance from the surface. As the charge nears the surface, the forces rise and would approach infinity as x approaches zero if the law were applicable at all distances. However, from the very nature of its derivation, which assumes a plane surface, it is obvious that this law will not hold for distances of less than a few atom diameters. The law which most nearly satisfies the observed facts assumes that the attractive force is zero at the metal surface, rises rapidly to a maxmium, then decreases inversely with the square of the distance

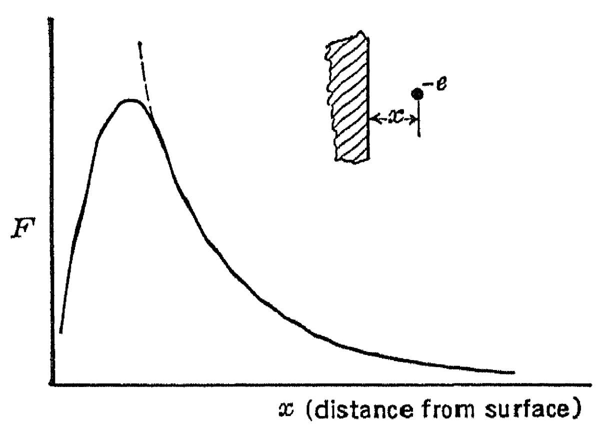


Fig. 1.1—Modified Image Force.

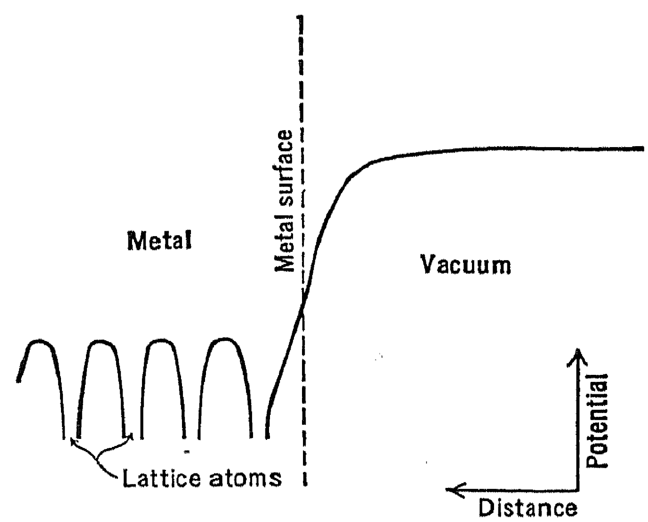


Fig. 1.2.—Idealized Potential Field.

from the metal surface. The force law close to the surface is extremely complicated, but for the present purpose a knowledge of its exact form is unnecessary. Fig. 1.1 shows the approximate shape of the force curve.

The potential at any point outside of the metal is calculated by integrating the force curve. The form of the potential curve is illustrated in Fig. 1.2. Inside the metal the potential is not uniform but consists of a three-dimensional array of potential minima, the minima corresponding to the crystal atoms. These minima are shown to the left of the metal surface in Fig. 1.2. For purposes of this discussion the non-uniformity of the potential within the metal can be ignored and the variable potential be replaced by a uniform averaged potential as shown in Fig. 1.3. The value of the potential rise at the surface of a metal is extremely important in determining its emission properties. This potential is known as the "inner potential" and is shown as W_a on the diagram. It can be measured experimentally by the change in wavelength of electron waves passing through the metal surface observed in electron diffraction experiments.

The values obtained by this method are in the neighborhood of 10 to 15 volts.

With these values for the inner potential, it is possible to calculate the approximate minimum distance at which the square law image force on an electron applies. This distance turns out to be of the same order of magnitude as the distance between atoms in the metal lattice.

For convenience of illustration, the kinetic energies of the electrons can be included in an energy diagram, together with the potential energy. The kinetic energy is plotted along the ordinate just as is the potential energy of the lattice. Such a diagram is shown in Fig. 1.3. The energy distribution in the electron gas, based upon the exclusion principle, was

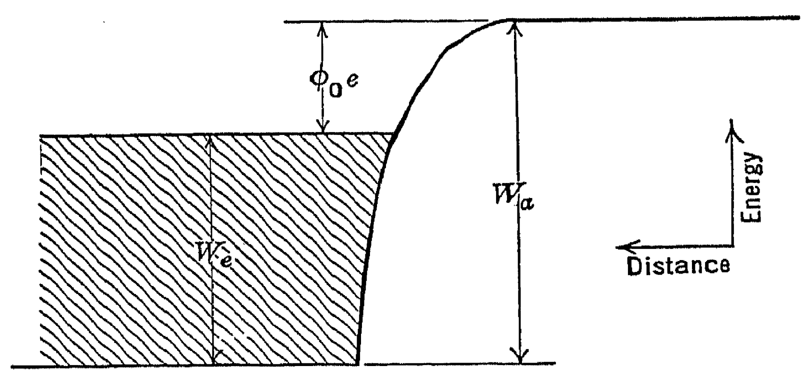


FIG. 1.3.—Electron Energies in an Ideal Metal.

derived by both Fermi and Dirac, independently, in 1926. According to their distribution function the number of electrons $N(u,v,w)dudvdw$ having velocity components u, v, w, in the range du, dv, dw is:

$$N(u,v,w)dudvdw = \frac{\frac{2m^3}{h^3}\, dudvdw}{\epsilon^{\frac{(W_e-\mu)}{kT}} + 1}, \tag{1.3}$$

where

$$\mu = \frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3}, \tag{1.4}$$

k is the Boltzmann constant, and
n is the number of electrons per unit volume.

A plot of this function is shown in Fig. 1.4. Curves are given for $T = 0$ and $T = 1600°$. At $T = 0$ this function becomes:

$$N(u,v,w) = 2\frac{m^3}{h^3} \qquad \text{for } W_e < \mu,$$

$$N(u,v,w) = 0 \qquad \text{for } W_e > \mu.$$

Thus there are no electrons in energy levels above $W_e = \mu$, which is in accordance with the qualitative picture given in the preceding paragraphs.

The numerical value for $W_{e\ \max.}$ can be found from the equation:

$$W_{e\ \max.} = \frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3} \tag{1.5}$$

However, to evaluate the expression it is necessary to assign a value to n. This requires a knowledge of the number of atoms per unit volume and the number of electrons each atom contributes to the electron gas. If the former is calculated from the density, composition, and Avogadro's number, and a reasonable assumption is made for the latter, $W_{e\ \max.}$ is found to lie between 5 and 10 volts.

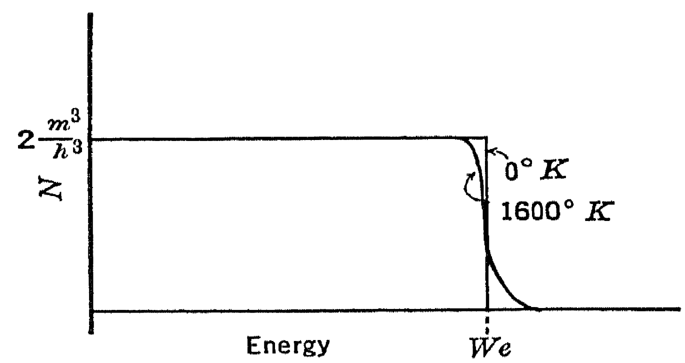


Fig. 1.4.—Distribution of Electron Energies by Fermi-Dirac Statistics.

At temperatures other than zero the velocity distribution does not end sharply at the value $W_{e\ \max.}$ but decreases asymptotically to zero. As the temperature is increased the high-velocity electrons become more and more numerous. Finally, at very high temperatures the Fermi-Dirac distribution approaches a Maxwellian distribution. However, this is found, as has been mentioned, to be a higher temperature than can be reached using any known metal.

When a potential difference is established between two points in a metal, electrical conduction takes place because the field thus produced causes a drift to be superimposed upon the random motion of the electrons in the upper levels. This drift is in a direction opposite to that of the field because of the negative charge of the electron. In other words, the drift motion is opposite to the conventional direction of current flow. Although the electrons behave largely as though they were free, there is some interaction between them and the crystal lattice. This interaction results in a loss of energy by the electrons to the lattice, which appears as the heat generated due to electrical resistance. In order to account quantitatively for electrical resistance of a metal it is necessary to deal

with the wave functions of the electrons and to consider resistance as a scattering and diffraction problem. Looked at in this way, both resistance and change of resistance with temperature can be accounted for very nicely.

1.2. Electron Emission. The new theory also applies to the emission of electrons. On the basis of the quantum-theory picture of metals nearly every phase of electron emission from pure metals can be explained not only qualitatively but also quantitatively. The theory, however, has not been extended to cover with the same completeness the case of metal surfaces on which there is a thin, overlying, contaminating film. Since

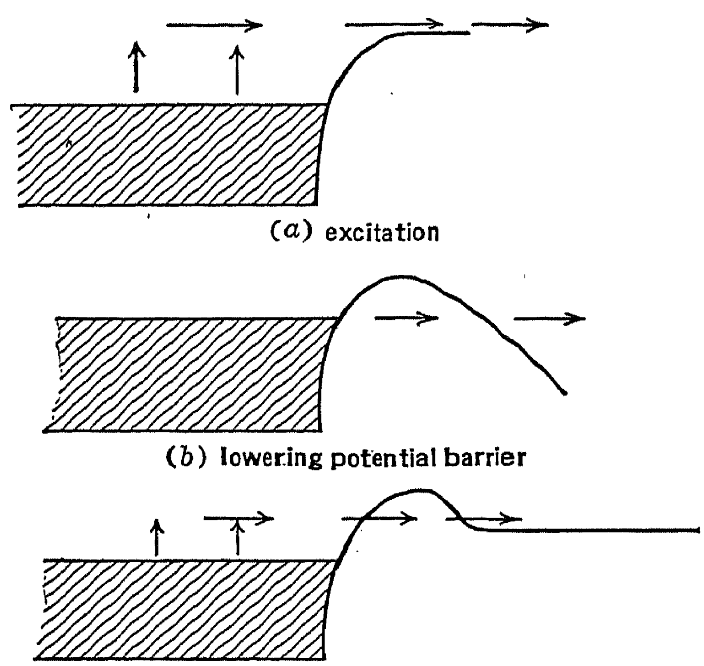


Fig. 1.5.—Emission of Electrons.

this type of surface is very important from the standpoint of emission of electrons and cannot be left out of consideration, there must necessarily be gaps and indefiniteness in the account that follows.

Turning again to Fig. 1.3 which shows the kinetic energies of the electrons superimposed upon the potential energy diagram for the metal, it is evident that one or both of two conditions must be fulfilled before electrons can escape through the surface barrier and be emitted. Either some of the electrons must be given additional energy so that their kinetic energy is at least as great as the potential barrier, or the height of the barrier must be reduced to such an extent that the electrons can penetrate through it. Diagrammatically, these might be represented as shown in Fig. 1.5. The first diagram illustrates electrons which have been given sufficient additional kinetic energy to escape over the normal potential barrier. This energy may be supplied in the form of thermal energy taken

from the lattice of a heated metal crystal, in which case the emission is known as thermionic emission. On the other hand, if the additional energy is obtained from radiant energy incident upon the metal, photoelectric emission takes place. Where the metal is bombarded with high-velocity electrons there may be an interaction between the electrons in the metal and the incident electrons, resulting in the former acquiring sufficient energy to escape, this phenomenon being known as secondary emission.

Diagram (*b*) shows the potential barrier depressed to such an extent that electrons can escape from the surface. This may occur when there are extremely high fields at the metal surface, the effect being known as cold emission. It should be pointed out that, while, according to classical theory, it is necessary to lower the potential barrier until it is less than the kinetic energy of the electrons, this is not so in quantum mechanics where there is a finite probability of an electron traversing a potential energy barrier which is higher than its kinetic energy. The probability of transmission here depends upon the thickness and the height of the barrier.

Very frequently there is both an addition of energy and a depression of the barrier, as shown in diagram (*c*). The depression of the barrier, may, for example, be caused by certain contaminations on the surface of the metal. When this occurs much less energy has to be added to the electrons to cause them to escape, thus making the substance a much better thermionic, photoelectric, or secondary emitting surface. This effect will be the subject of much of the discussion relating to practical emitters.

1.3. Shot Effect. As has been stated, energy is supplied to some of the vast numbers of free electrons in order to produce thermionic emission, photoelectric emission, etc. The fraction of electrons which receive enough energy to permit them to escape is extremely small, and only part of these move in such a direction that they are emitted. The whole process is, on this basis, an entirely random phenomenon, each electron being emitted entirely independently of any other.

Since it is a random process, the number of electrons emitted during any given interval differs by a small amount from the average number of electrons emitted in an interval of this duration. The electron current from an emitter is therefore not constant, but is subject to statistical fluctuation. As this fluctuation limits the performance of many types of electronic devices, it is important to determine its magnitude and nature.

An emitter which releases N electrons in a time T, where N is a large number, has an average rate of emission of $n_0 = N/T$ electrons per unit time. However, since the emission phenomenon is a random one, the

actual number n emitted during any given unit of time will differ from n_0. According to the laws of probability, while the average difference:

$$\overline{(n - n_0)}$$

must equal zero, the mean square difference over the period T is not zero but rather is given by:

$$\overline{(n - n_0)^2} = \overline{\Delta n^2} = n_0. \tag{1.6}$$

Since the emission of each electron corresponds to the transfer of a charge $-e$, Eq. 1.6 can be expressed in terms of charge, giving

$$\begin{aligned} \overline{(\Delta ne)^2} &= e^2 n_0, \\ \overline{\Delta q^2} &= Qe. \end{aligned} \tag{1.7}$$

In other words, the mean square fluctuation of charge is proportional to the charge emitted.

In the form given above in Eq. 1.7 the relation is not very directly applicable to actual electrical circuits. In general, the circuits encountered respond to frequencies which lie in a band between definite limiting frequencies. The useful signal is in the form of current variations over the frequency band in question. The shot effect, or statistical fluctuations, introduces current variations over this same frequency band which tend to mask the useful signal. Therefore Eqs. 1.6 or 1.7 must be reformulated to express the magnitude of this unwanted random "noise" in terms of a fluctuating current in the circuit containing the emitting element.

Such a relation was first worked out by Schottky in 1912. The derivation given below,* though not identical with that originally given by Schottky, yields the same information. As a starting point, the current pulse resulting from the motion of an individual electron is expressed as a Fourier integral which gives its frequency spectrum. The mean square of a large number of these current pulses is then found by summing and averaging.

The current arising from a single emitted electron can be expressed as follows:

$$i_1(t) = \frac{e}{\pi}\int_0^\infty d\omega \cos \omega\,(t - T_1), \tag{1.8}$$

where T_1 is the time at which the electron is actually emitted. This equation assumes that the emitted electron produces a single current pulse whose duration is extremely small compared with any interval of time for which Eq. 1.8 is to be used.

* From the unpublished work of E. G. Ramberg.

The mean square current due to N electrons emitted in the time T is therefore:

$$\overline{i^2} = \frac{1}{T}\int_0^T dt\left[\sum_{v=1}^{N} \frac{e}{\pi}\int_0^\infty d\omega \cos \omega\,(t - T_v)\right]^2. \tag{1.9}$$

Carrying out the indicated squaring of the sum over the N electrons, this equation becomes:

$$\overline{i^2} = \frac{e^2}{\pi^2}\frac{1}{T}\int_0^T dt\int_0^\infty d\omega\int_0^\infty d\omega' \sum_{v=1}^{N}\sum_{v'=1}^{N} \cos \omega\,(t - T_v) \cos \omega'\,(t - T_{v'}).$$

This integral expression must be integrated over ω' and t. When performed, this integration leads to the equation:

$$\overline{i^2} = \frac{e^2}{\pi^2}\frac{1}{2T}\int_0^\infty d\omega \sum_{v=1}^{N}\sum_{v'=1}^{N} 2\pi \cos \omega\,(T_v - T_{v'}), \tag{1.10}$$

which expresses the mean square current over an infinite frequency band.

Since the response of any real circuit is zero except over one or more finite bands of frequency, an expression for the current fluctuation in a limited band covering the frequencies between f_1 and f_2 is of value. For this the limits of integration over ω instead of being 0 to ∞ are taken as $2\pi f_1$ and $2\pi f_2$. Eq. 1.10 consequently becomes:

$$\begin{aligned}\overline{i^2} &= \frac{e^2}{\pi^2}\frac{1}{2T}\int_{2\pi f_1}^{2\pi f_2} d\omega \sum_{v=1}^{N}\sum_{v'=1}^{N} 2\pi \cos \omega\,(T_v - T_{v'}) \\ &= 2\,\frac{e^2N}{T}\,(f_2 - f_1) + \frac{e^2}{\pi T}\sum_{\substack{v=1 \\ v\neq v'}}^{N}\sum_{v'=1}^{N} \frac{\sin 2\pi f_2(T_v - T_{v'}) - \sin 2\pi f_1(T_v - T_{v'})}{T_v - T_{v'}}.\end{aligned}$$

The second term of this expression vanishes because of the assumed randomness of the emission. That is, since the time interval $T_v - T_{v'}$ between the emission of any two electrons is purely random, the summations represent the superposition of a vast number of sines of random angles which are as likely to have a positive as a negative value, and therefore average zero. Thus the mean square fluctuation current over the frequency band $f_2 - f_1$, for an average emission current i_0 ($i_0 = Ne/T$), reduces to:

$$\overline{i^2} = 2ei_0(f_2 - f_1). \tag{1.11}$$

As is evident from the general nature of the derivation, Eq. 1.11 applies to nearly all types of electron emission. Actually the relation accurately gives the fluctuations in all forms of photoelectric and thermionic

emission and in cold emission if the voltage is well below the rupture point. Secondary emission is an exception, because each incident electron causes the emission of several electrons, giving a form of coherence between the emitted electrons which does not fulfill the condition of random emission upon which the derivation is based.

The statistical current fluctuation given by Eq. 1.11 represents the minimum that can be expected from any emitting surface in the absence of space charge. Other effects not directly related to the basic phenomenon of electron emission may operate to increase the total fluctuation of the current emitted.

The following problem of a phototube emitting electrons in response to a modulated light signal serves to illustrate a typical noise calculation. Let it be assumed that the photoelectric current is amplified by an amplifier which responds to a frequency band F, 10,000 cycles wide. If the photocurrent is $i_0 = I\ (1 + k \sin 2\pi ft)$, where k is a modulation factor which has a value 0.1, what is the minimum value of I which will give a signal-to-noise ratio of 10?

In this case the mean square signal is: $\overline{i_s^2} = \frac{1}{2}k^2I^2$ and the mean square noise is $\overline{i_n^2} = 2\ eFI$.

The signal-to-noise ratio is therefore:

$$R = \sqrt{\frac{\overline{i_s^2}}{\overline{i_n^2}}} = \frac{\frac{1}{\sqrt{2}}kI}{\sqrt{2eFI}},$$

$$I = \frac{R^2 2eF}{\frac{1}{2}k^2};$$

or, in the example in question:

$$I = \frac{100 \times 32 \times 10^{-20} \times 10^4}{\frac{1}{2} \times 10^{-2}}$$

$$= 0.64 \times 10^{-10} \text{ ampere.}$$

In the foregoing it is assumed that the entire noise was the shot noise of the emitted electrons in the phototube. Actually noise is generated in the resistor used to couple the phototube to the grid of the first tube of the amplifier, and noise is also introduced by the first tube itself. The effects of these elements will be discussed in Chapters 10 and 14, in which problems of amplification are considered.

1.4. Thermionic Emission. The general properties of the emission of electrons from metals having been considered, it is now expedient to treat the different types of emission individually. As has already been stated, it is possible to cause electrons to be emitted by heating the metal, this effect being known as thermionic emission. The emission current obtained thus will depend upon the height of the potential barrier above the top of the Fermi distribution of free electrons, upon the shape of the barrier, and upon the energy distribution of the electrons.

The energy distribution can be treated quite generally using the Fermi-Dirac distribution function, which is applicable to all metals. The height and shape of the potential barrier, however, depend not only upon the particular metal, but also upon the state of the surface of the emitter. Three classes of emitters having different types of potential barriers will be considered. These are: emitters having a pure metal surface, those having an absorbed electropositive layer, and finally the more complicated oxide-coated cathodes.

To calculate the electron emission it is necessary in the first place to determine the number of electrons arriving in unit time at the surface of the emitter with sufficient kinetic energy directed towards the surface to pass through the potential barrier. This can be done with the aid of Eq. 1.3, as was shown by Nordheim, who derived the thermionic emission equations on a quantum mechanical basis. The Fermi function is integrated over the velocity components v and w taken from zero to infinity. This expression when rewritten to express the number of electrons $N(W)dW$ which have a kinetic energy in the range W to $W + dW$ associated with the component of velocity u which is assumed to be directed normal to the metal surface gives the following distribution function:

$$N(W)dW = \frac{\pi}{h^3}\sqrt{\frac{8m^3}{W}}\, kT \ln\left(1 + \epsilon^{-\frac{W-\mu}{kT}}\right)dW. \tag{1.12}$$

Then, since only those electrons having an energy W greater than W_a can escape through the boundary, the total number of electrons leaving a unit area per unit of time will be:

$$n_0 = \int_{W_a}^{\infty} \sqrt{\frac{2W}{m}} N(W)dW. \tag{1.12a}$$

For convenience in performing the indicated integration, the distribution function $N(W)$ can be expressed as:

$$\sqrt{\frac{2W}{m}}\, N(W)dW = \frac{4\pi mkT}{h^3}\, \epsilon^{-\frac{W-\mu}{kT}} dW, \tag{1.12b}$$

since over the range of integration $\epsilon^{-(W-\mu)/kT}$ is small. Eq. 1.12a therefore becomes:

$$n_0 = \int_{W_a}^{\infty} \frac{4\pi mkT}{h^3} \epsilon^{-\frac{W-\mu}{kT}} dW$$
$$= \frac{4\pi mk^2}{h^3} T^2 \epsilon^{-\frac{W_a-\mu}{kT}}. \tag{1.13}$$

Experimentally and theoretically it can be shown that some of the electrons in the energy range above W_a are reflected back. If r is taken as the reflection coefficient, the fraction of electrons escaping is $(1 - r)$.

Including this transmission coefficient, combining the factor $4\pi mk^2e/h^3$ into a single constant A_0, and expressing the rate of emission of electrons in terms of current, Eq. 1.13 can be written as:

$$i_0 = (1 - r)A_0T^2\epsilon^{-\frac{W_a-\mu}{kT}}. \tag{1.14}$$

It should be noticed that this equation has practically the same form as the thermionic emission equations derived by Richardson, Dushman, and others on the basis of classical thermodynamics.

In the exponent of this equation there appears the term $(W_a - \mu)$. This represents the energy difference between the height of the potential barrier and the maximum kinetic energy of the electrons in the Fermi distribution at zero absolute, and is given the name work function. It is common practice to express this work function in terms of electron volts, so that the exponent in emission equations takes the form $-\phi_0 e/kT$, or to simplify it still further by the substitution $b_0 = \phi_0 e/k$. Also, for convenience, the coefficient A_0 and the reflection factor are often combined into a single constant A. When thus written, Eq. 1.14 becomes:

$$i = AT^2\epsilon^{\frac{-b_0}{T}}. \tag{1.14a}$$

A few representative values for the constants A and b_0 and the work function ϕ_0 are given in Table 1.1.

The thermionic emission from the metals listed is very small when the metal is heated to any reasonable temperature. For example, tungsten heated to 2200°K will emit only 13 milliamperes per square centimeter. This is because for pure metal surfaces the work function is high. Calcium or caesium, whose work functions are low, have such low melting points that they cannot be used as thermionic cathodes.

The most completely studied composite surface thermionic emitter is that consisting of thorium on tungsten. From Table 1.1 the work function of tungsten alone is 4.52 volts while that of thorium is 3.35 volts. How-

TABLE 1.1 *

Element	ϕ_0	$b_0 \times 10^{-3}$	A
Calcium	2.24	26	60.2
Caesium	1.81	21	162
Columbium	3.96	46	57
Molybdenum	4.44	51.5	60.2
Nickel	2.77	32.1	26.8
Platinum	6.27	72.5	1.7×10^{-4}
Tantalum	4.07	47.2	60.2
Thorium	3.35	38.9	60.2
Tungsten	4.52	52.4	60.2
Zirconium	4.13	47.9	330

* From L. R. Koller, reference 7.

ever, a surface consisting of tungsten having over it a monatomic layer of thorium covering about 70 per cent of its surface has an effective work function of only 2.6 volts.

If a contaminating layer of caesium is used instead of thorium the work function is still further depressed, being only 1.6 volts. Unfortunately this surface is unstable and the caesium evaporates from the base metal before temperatures can be reached which will give a high electron yield. Unlike the caesium surface, the thoriated surface is stable up to temperatures of 2000 to 2200°K or above, which property makes this emitter very valuable in practical electronics. The emission from a thoriated tungsten cathode at 2000° is more than 1000 times that from pure tungsten.

The explanation of the increased emission of these complex surfaces lies in the nature of the work function. The work function may be divided into two parts, the first being due to the attraction of the lattice ions as a whole for the electrons, and the second to the dipole moment of the surface. The first remains unchanged when a contaminating layer is added to the surface; the second factor may be greatly altered in either direction. When a caesium atom approaches a tungsten surface it enters a region of fairly high positive potential which tends to remove its valence electron, and since the work function of tungsten is greater than the ionization potential of the caesium, the atom will be actually ionized. The caesium thus adheres to the surface in the form of a positive ion. As more caesium atoms arrive, these tend to ionize also. However, the addition of each ion tends to lower the work function of the surface, so that after a certain number have adhered the work function is no longer

sufficient to produce ionization. Even under these conditions partial ionization or polarization takes place. A tungsten surface, therefore, which is more or less completely covered with caesium may be thought of as being coated with a layer of caesium "partial ions" which cause a lowering of the effective work function ϕ_0.

In the case of thorium, the work function of the tungsten is less than the ionization potential. This does not preclude the existence of ions on the surface, since both image force and work function act to produce ionization. However, unlike elements which are ionized by the work function alone, thorium evaporated from the surface does not leave in the form of ions. Beyond a certain coverage "partial ions" only are formed. Both thorium ions and partial ions contribute to a lowering of the work function.

If, instead of an electropositive contaminating atom, an electronegative element is allowed to form a surface layer on the tungsten, the opposite effect occurs and the work function is raised. For example, a monatomic layer of oxygen on tungsten raises the work function to 9.5 volts.

A more complex layer can be formed if oxidized tungsten is exposed to caesium vapor, thus allowing the formation of a final positive layer. In this case the work function is lower than for caesium on pure tungsten and a more stable surface is formed. Such a surface will yield 350 milliamperes per square centimeter at 1000°K.

In view of the practical importance of the thoriated tungsten emitter, it will not be out of place to discuss in greater detail the actual form of these cathodes.

The base is metallic tungsten to which has been added 0.5 to 2.0 per cent by weight of thorium oxide. The material is shaped to the form required for the cathode, assembled in the electron tube in which it is to be used, and the tube exhausted. To activate the emitter it is first flashed for a few seconds at 2800°K, then aged at 2100°K for a matter of several minutes. The cathode is then ready for operation.

At the flash temperature some of the thoria reduces to metallic thorium. This metallic thorium diffuses through the base material to the surface. Electron microscope examination shows that this diffusion is partly in the form of minute eruptions through the crystal grains, and partly along grain boundaries. At temperatures as high as 2800°K it is evaporated from the surface immediately upon its arrival. In other words, the rate of evaporation greatly exceeds the rate of diffusion.

As the temperature is decreased, both the rate of diffusion and the rate of evaporation decrease. However, the evaporation rate falls the faster with temperature. Below 2400°K a layer of thorium begins to

form on the surface. At the aging temperature of 2100°K an optimum coverage is obtained. At still lower temperatures the evaporation rate reaches an extremely low value; under this condition the cathode is stable and its operating life adequate. If the emitter loses its activation it can usually be restored by again flashing it both to clean possible poisonous contaminations from the surface and to reduce more thoria, and then aging it at 2100°.*

For circuits handling low power and for both television pickup and viewing tubes, the oxide-coated cathode is by far the most important. Theoretically, it is the least understood of the thermionic emitters.

This cathode consists of a thick layer of a mixture of barium and strontium on a nickel base. The best results have been obtained when the base metal contains a small amount of reducing agent such as Si, Ti, Al, etc. However, fairly high emission can be obtained from the oxide mixture on almost any base metal.

The material is prepared by coprecipitation of strontium carbonate and barium carbonate to form a mixture containing about equal parts by weight of the two components. The mixture, suspended in a suitable organic binder, is painted or sprayed on a nickel base, either before or after the base has been assembled in the final structure. After the tube has been pumped but before it is sealed off the vacuum system, the cathode is heated to approximately 1300°K for one minute. This drives off the binder and reduces the carbonates to the oxide. The temperature is then reduced to about 1100°K, and a voltage is applied between the cathode and the nearest element. The temperature being left constant, the voltage is gradually increased. It will be found that the current drawn to the electrode increases and then reaches saturation. With constant voltage and temperature the tube is aged for a period of 10-20 minutes. The cathode is then ready for use. The operating temperature of this type of cathode is from 1000 to 1050°K. An oxygen-bearing film may be formed on any electrode near the cathode which is exposed to the evaporation products of the cathode. If the electrode is subjected to electron bombardment, the film will decompose, releasing oxygen. This leads to de-activation of the cathode and falling emission. The formation of the film is prevented by maintaining the electrode at a temperature of 1250°K, or above.

This is only one of the many activation schedules used in practice, the

* In practical vacuum-tube work involving high voltages and high power, thoriated tungsten is not used because of the severe electrical conditions under which the cathode operates. Under these conditions pure tungsten has been used almost exclusively as the material for the emitter. Recent measurements on the rate of evaporation indicate that tantalum may be a suitable electron source.

particular activation found most satisfactory depending upon the specific device in which the cathode is to be used. These variations do not, in general, increase the specific emission from the cathode material but often decrease the percentage of failures.

As has been stated, the exact mechanism of the emission from this type of cathode is not clearly understood. During the activation some of the barium oxide is reduced to metallic barium either by reaction with the base metal, thermal decomposition, or electrolysis. Part of this barium forms a surface layer on the cathode, this being apparently a necessary condition for electron emission. The percentage of barium oxide mixed

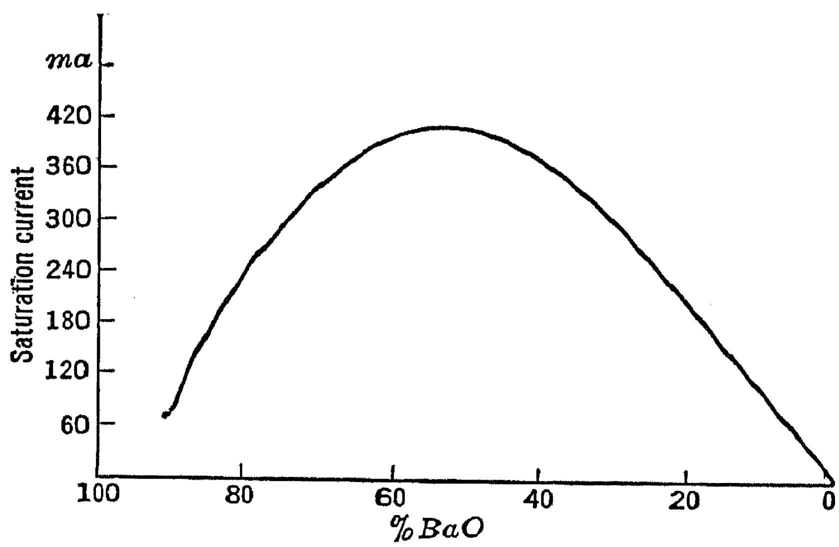


Fig. 1.6.—Variation of Saturated Emission from an Oxide Cathode with Percentage of Barium Oxide.

with the strontium oxide is not very critical. Fig. 1.6 shows the relation between the barium oxide strontium oxide ratio and the emission.

A properly activated oxide-coated cathode has an effective work function of about 1 volt and a working emission of 300 to 400 milliamperes per square centimeter at 1000°K. This type of emitter lends itself well to the indirectly heated unipotential type of cathode. Two modifications are shown in Fig. 1.7. In this diagram (*a*) shows an indirectly heated cathode typical of those used in the ordinary receiving tube, and (*b*) shows the type of emitter used in the television pickup and viewing tube. Cathodes of this type are eminently suited for use where alternating current is supplied to the heater; they have a long life and will stand a great deal of abuse without losing their activation.

Before leaving the subject of thermionic emission, mention must be made of the variation of current with applied voltage in a normal two-electrode vacuum tube. The current-voltage relation for a diode consist-

ing of a thermionic emitter and a collecting plate can be represented by the curve shown in Fig. 1.8. When the plate is slightly negative a few electrons reach it owing to the fact that some of the electrons have a fairly high initial velocity, the number having any given velocity being determined by the Maxwellian velocity distribution of emitted electrons.

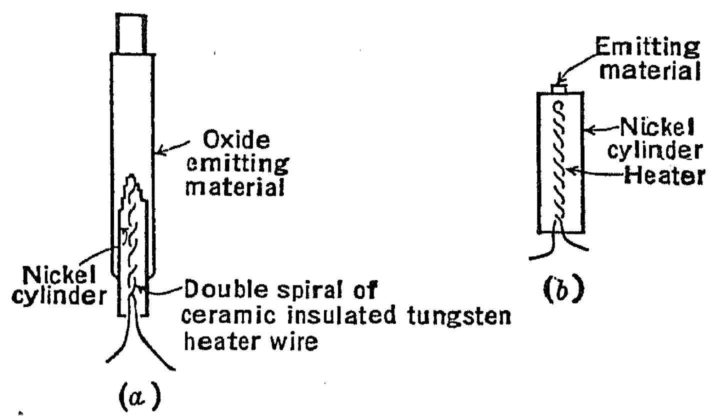


FIG. 1.7.—Indirectly Heated Cathodes.

As the voltage between the cathode and plate is reduced to zero, more electrons reach the plate. However, the number arriving will be less than the maximum emission, owing to the retarding effect of the cloud of electrons lying between the two elements. This is known as the space-charge effect. The action of space charge can be most readily seen by considering

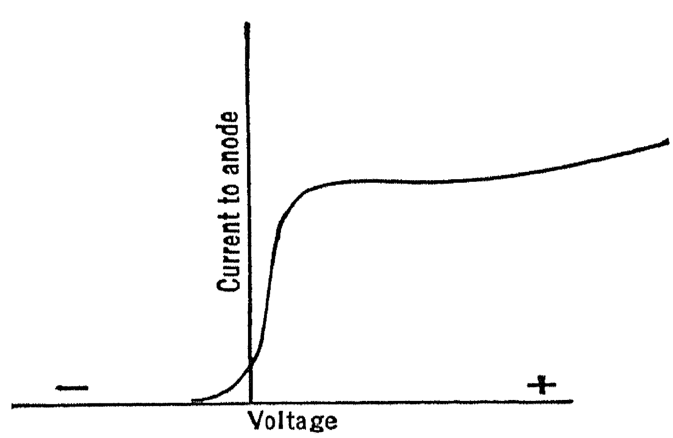


FIG. 1.8.—Idealized Current Characteristics of a Diode.

the potential distribution between the cathode and plate. This potential is determined not only by the voltage applied to the electrodes, but also by the charge between them, and to determine the potential distribution it is necessary to integrate the Poisson equation $\Delta\phi = 4\pi\rho$, where ρ is the density of charge in the intervening space. Fig. 1.9 shows a typical

family of potential-distribution curves for various voltages between emitter and plate. Clearly, to obtain the total emission current, it is necessary to apply a positive potential such that a potential minimum does not exist between the two electrodes. The initial rise in current with increasing positive potential on the plate, shown in Fig. 1.8, is the result of overcoming the effect of space charge; over this portion of the curve the current is proportional to $\phi_0^{3/2}$. The rise continues until the potential minimum between plate and cathode no longer exists. When this condition is reached the current is practically independent of voltage. This value of current is called the saturation current.

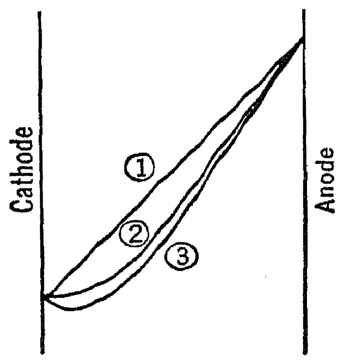


FIG. 1.9.—Effect of Space Charge on Potential Distribution.

Actually, the current continues to increase as the voltage is further raised. This effect is due to the lowering of the potential barrier by the external field. The relation between the applied voltage and the collected current is:

$$i = i_s \epsilon^{\frac{4.398\, E^{\frac{1}{2}}}{T}}, \tag{1.15}$$

where i_s is the saturation current and E the field strength at the emitter in volts per centimeter. This formula was calculated by Schottky for pure metal surfaces. Because of this contribution the phenomenon has become known as the "Schottky effect."

This relation does not apply to composite surface emitters or to oxide-coated cathodes. Such emitters do not show a true saturation effect, but instead merely a change in the rate of rise of current at the point where space charge no longer limits the current flow. Beyond this point the current increases much more rapidly with increasing field than is indicated by the Schottky equation.

1.5. Photoelectric Emission. Photoemission, i.e., the emission of electrons from matter by radiant energy, is of fundamental importance in television. In order to account for the experimental facts observed in connection with this phenomenon it is necessary to make use of the quantum theory of radiation, as well as the quantum mechanical picture of metals. On the basis of this theory, radiation must be considered corpuscular in nature, when dealing with any problems involving energy interchange. Accordingly, radiation of frequency ν behaves as though it were composed of a vast number of particles or photons, each with energy $h\nu$. The energy in each individual photon is independent of the intensity of the radiation of which it is a part, and depends only upon

its frequency. When a photon interacts with an electron, giving up its energy, it can increase the kinetic energy of the electron by an amount equal to or less than $h\nu$.

When radiation falls upon a clean metal surface which is assumed to be at such a temperature that very few of its electrons have an energy greater than that corresponding to the top of the filled Fermi band, some of the photons will give up their energy to free electrons. Since the maximum energy any of these electrons can have before collision with a photon is $W_{e\ \mathrm{max.}}$, the maximum that they can have as the result of a collision is $W_{e\ \mathrm{max.}} + h\nu$. If this energy is greater than the height of the barrier W_a, those electrons which happen to be moving toward the surface can escape. Thus, to obtain photoelectric emission, it is necessary that the frequency of the incident radiation be greater than the work function $e\phi_0$ divided by Planck's constant h. When the frequency is in excess of this value, the photoelectrons will be emitted with an initial velocity corresponding to the difference in energy. The maximum initial velocity V_e in electron volts will therefore be given by the relation:

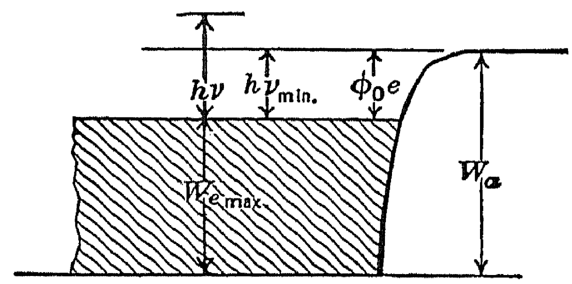


FIG. 1.10.—Electron Energies Associated with Photoelectric Emission.

$$V_e = \frac{h\nu}{e} - \phi_0. \tag{1.16}$$

This relation has been tested, by retardation methods, for a great many pure metal surfaces, and is found to be fulfilled within the accuracy of the assumptions made in deriving the expression. Chief among the assumptions are those that no electrons have energies greater than $W_{e\ \mathrm{max.}}$ and that there are no multiple collisions between electrons and photons. Actually, electrons are emitted with velocities greater than is indicated by Eq. 1.16 owing to the existence in the metal of electrons with energies in excess of $W_{e\ \mathrm{max.}}$. More exact values for the photoelectric work function are obtained by making use of the relation derived by Fowler giving the photoelectric yield in terms of the retarding potential and temperature.

The wavelength of the radiation for which $h\nu = e\phi_0$ is known as the photoelectric threshold, or long-wavelength limit, and is characteristic of the surface. Table 1.2 gives the long-wavelength limit for a number of surfaces, the work function as determined by this threshold, and the thermionic work function.

TABLE 1.2*

Metal	Long-Wavelength Limit	Photoelectric Work Function	Thermionic Work Function
Ag	2610Å	4.73	4.08
Al	3500 (approx.)	2.5 to 3.6	—
Au	2650	4.82	4.42
Ca	(4500)	(2.7)	2.24
Cs	6600	1.9	1.81
K	5500	1.76 to 2.25	—
Mg	>3650	<3.4	—
Pt	1962	6.3	6.27
W	2650	4.58	4.52
Zn	3720	3.32	—

* Data from Hughes and DuBridge, reference 6.

Like thermionic emission, the photoelectric emission is greatly increased and the long-wavelength limit raised when a metal surface is subjected to certain contaminations. This is particularly true when the superimposed layer is an alkali metal. Surfaces having the greatest photoelectric response are obtained when complex thick surface layers, having as a final coating an alkali metal, are used. These will be described in detail later.

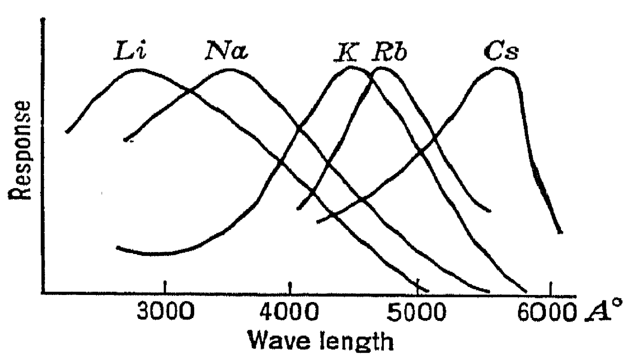


Fig. 1.11.—Spectral Selective Effect of the Alkali Metals (after Hughes and DuBridge).

The alkali metals, together with a few others, and nearly all composite surface emitters, exhibit a property known as selective photoemission. This effect can be subdivided into polarization selectivity and spectral selectivity. The two types of selectivity have been found occurring together wherever observations have been made, but the investigation has not been carried far enough to prove with certainty that they have a common origin. Photoemission, when plotted as a function of frequency, gives a curve falling smoothly to zero at the long-wave threshold in the case of most clean metal surfaces. However, when spectral selectivity is present, the emission curve is not smooth but has one or more maxima which rise to many times the height that would be expected without selective emission. Fig. 1.11 illustrates the spectral selectivity of the alkali metals.

When polarized light is used for this measurement, it is found that the response to light polarized so as to have its electric vector parallel to the emitting surface is much smaller than for light polarized in the other plane. The ratio of the response for the two directions of polarization is illustrated in Fig. 1.12, for several angles of incidence. It is seen to increase rapidly with the size of the angle.

The surface of the ordinary practical photosensitive emitter is so rough that no differentiation can be made between the two planes of polarization, and measurements yield an averaged spectral selectivity.

The problem of obtaining a high photoelectric yield is twofold. First, the work function of the surface must be made low, to increase the probability that a given electron-photon collision will give the electron at least sufficient energy to cross the potential barrier. Secondly, the quantum efficiency must be high, that is, the probability of an energy interchange between a photon and an electron should be as great as possible.

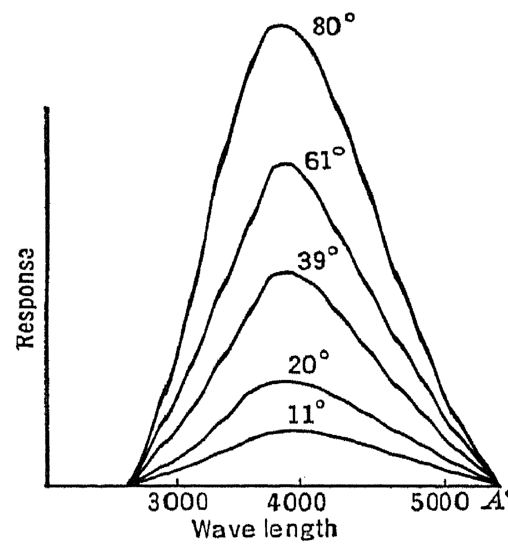


Fig. 1.12.—Dependence of Selective Photoelectric Emission on Angle of Incidence for Na-K Alloy (after Hughes and DuBridge).

As a consequence of the first condition, all efficient photoemitters are complex surfaces similar to those discussed under thermionic emitters. Owing to the difference in quantum efficiencies, however, there is not an exact correspondence between the efficiency of a surface as a thermionic emitter and as a photoelectric emitter. There exist at present considerable empirical data bearing on this phase, but only a meager theoretical understanding. Hence it must suffice to describe a number of surfaces suitable for various spectral ranges without attempting to account for their behavior.

The photosensitive surface most widely used today is the so-called caesiated silver surface. This type of emitter is extremely sensitive in the red end of the visible spectrum, and its sensitivity extends well down into the infra-red, the long-wave threshold for an ordinary surface being in the neighborhood of 12,000 or 13,000Å, while special emitters of this type have been made which extend as far as 17,000Å. The emission is selective, having a maximum just at the red end of the visible spectrum, a minimum in the green, and then an increase in the ultra-violet region. Fig. 1.13 shows a typical spectral response curve for this type of surface.

The spectral response is sensitive to the exact schedule of activation

used in preparing the emitter, and response curves which differ greatly from that given in Fig. 1.13 can be obtained.

The preparation of this type of surface is, briefly, as follows: A silver, or silver-coated, sheet of metal is shaped to the form required for the cathode in question, and mounted. Often the silver is etched to give a matte surface. It is next thoroughly cleaned, both with solvents to remove all traces of organic materials, and by heating in vacuum. After the cleaning process has been completed, the silver is oxidized. This is done by admitting oxygen at a pressure of a few millimeters of mercury into the tube containing the cathode and passing a glow discharge between the cathode and any other convenient electrode. As the surface oxidizes it will be found to go through a series of colors in the following order: yellow, red, blue, yellow, red, blue-green (or second blue), green.

In general the oxidation should be stopped with the second blue or first green. Caesium is then admitted to the tube, and the tube is baked for a few minutes at around 200°C. If the correct amount of caesium has been added the color of the finished surface will be a brownish yellow. Too much caesium is indicated by an almost white appearance of the surface, while an insufficient amount of caesium will leave the surface a dark gray-green.

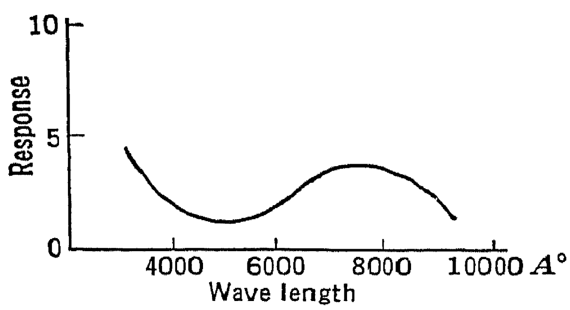


FIG. 1.13.—Spectral Response of Cs-CsO-Ag (Zworykin and Wilson).

The completed surface is thought to consist of a layer of caesium oxide interspersed with reduced silver particles and having a film (possibly in the form of discrete atoms) of caesium covering it.

The sensitivity of this type of emitter can be still further increased if a very thin layer of silver is evaporated over the surface of a cathode prepared as described above, and the emitter again baked at 200°C. This treatment increases the long-wavelength limit and also increases its sensitivity to white light.

A similar surface can be prepared with rubidium instead of caesium. This type of cathode has a selective spectral response which closely matches that of the eye, and for that reason is suitable for color work, such as, for example, the transmission of colored television pictures. Emitting surfaces can also be prepared from the other alkali metals—these having the peak of their response in the blue-violet or ultra-violet portions of the spectrum.

Caesium, in conjunction with such metals as antimony, bismuth, arsenic, alloys of antimony and silver, and many others, forms interesting and important photosensitive emitters. These are frequently of value

when a particular spectral response is wanted, as each combination has a more or less characteristic selectivity. For the ultra-violet portions of the spectrum, phototubes activated with potassium which has been subjected to a hydrogen discharge are frequently employed. Untreated uranium surfaces are also found to be very effective in the ultra-violet region. The advantage of these cathodes over caesiated silver photoemitters is that the dark current, owing to thermionic emission, ions, and similar causes, is much lower. For that reason these tubes are frequently advantageous where minute intensities of ultra-violet radiation are to be measured.

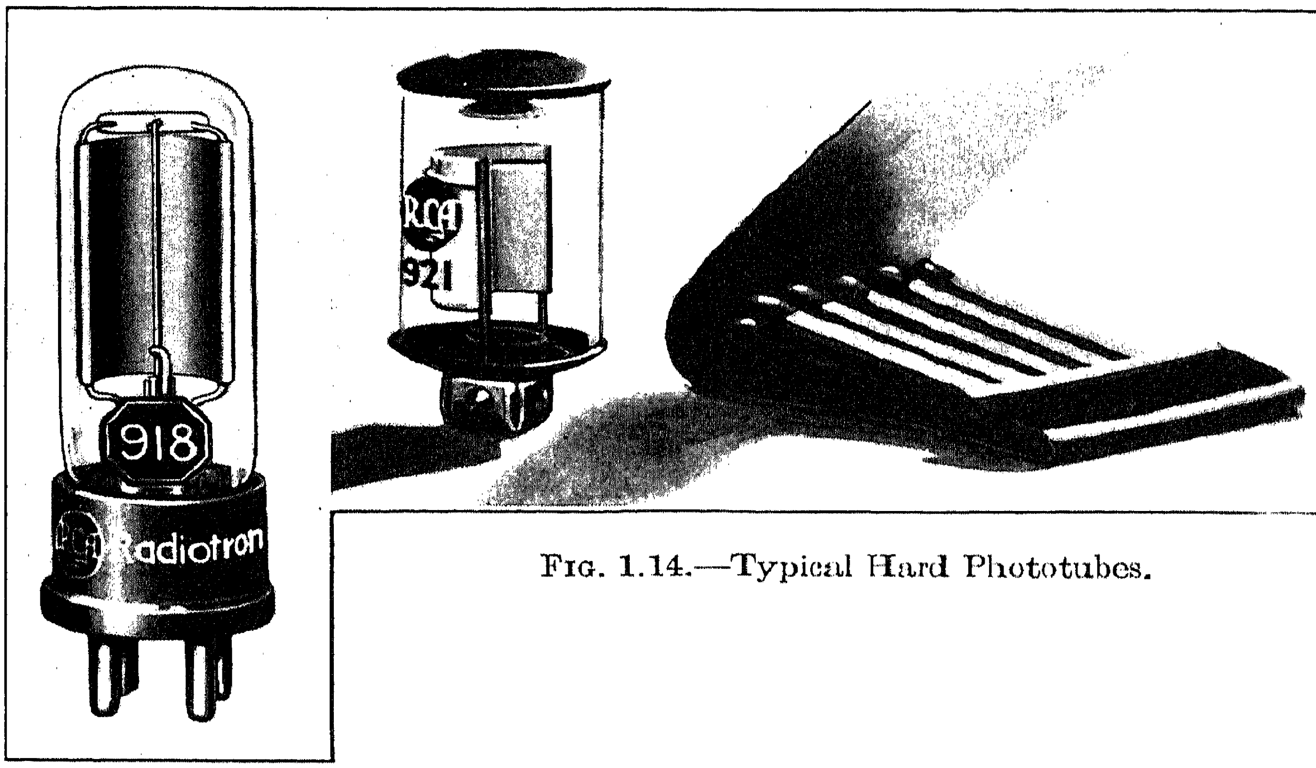


Fig. 1.14.—Typical Hard Phototubes.

Caesiated silver surfaces of the type described above are found in nearly all commercial phototubes. Fig. 1.14 is a photograph of two ordinary commercial hard phototubes. Such a tube will yield a photoelectric current of 20 to 50 microamperes per lumen (an incandescent tungsten filament heated to 2800°K serving as the light source). Phototubes of this type are evacuated to as low a pressure as possible.

Another commercial phototube employs a similar caesiated silver surface, but instead of being exhausted to a very low pressure the tube contains about 100 microns of an inert gas. When it is operated with about 90 volts between the anode and the cathode, the gas pressure is insufficient to allow a glow discharge. However, when electrons are released from the cathode they ionize some of the gas molecules in the course of their passage between cathode and anode. The ions thus generated pro-

duce electrons which enhance the current reaching the anode. In this way it is possible to multiply the original photocurrent by as much as 5 or 10 times. Such a gas phototube may have a photoelectric response which is as much as 100 to 200 microamperes per lumen. The disadvantages of this type of tube are twofold. First, the dark current is quite large, compared to that of a hard tube; and second, the gas amplification is a function of the frequency when the excitation is intermittent light, falling to a very low value at frequencies above 5000 cycles per second.

Before the question of photoemission is left, mention should be made of the semi-transparent cathode. This type of emitter is made in such a way that, when illuminated from the rear, photoelectrons are emitted from its face. Where it is desired to obtain an electron image of an extended optical image, such surfaces are important, as will be explained later.

The semi-transparent cathode is quite similar in nature to the caesiated silver surface described above, although the procedure of forming it is somewhat different. A thin film of silver is evaporated onto the transparent surface which is to support the cathode. This film should be of such a thickness that it transmits about 30 per cent of white light incident on it. It is then subjected to an oxygen discharge until it becomes almost completely transparent. Caesium is next introduced and the tube is baked. Like the ordinary caesiated silver cathode this surface is improved by evaporating a thin layer of silver over the activated film. A film prepared in this way will have a photosensitivity of 10 to 20 microamperes per lumen and a threshold at about 12,000Å.

1.6. Secondary Emission. A third way in which electrons can be given the energy necessary for their escape through the potential barrier is through the interaction between them and high-velocity electrons bombarding the surface of the emitter. The name secondary emission has been given to this phenomenon.

Secondary emission, like photoemission, is produced by an external excitation. It differs, however, in that each incident electron, unlike the incident photons which produce photoemission, may cause the release of more than one electron, the average number released by each primary electron being known as the secondary-emission ratio. This ratio may range from less than one electron per primary to more than ten or fifteen, depending upon the surface bombarded and the conditions of bombardment.

As was mentioned earlier in this chapter, the theoretical aspect of this phenomenon is far from clear, particularly in the case of the bombardment of complex surfaces. However, the observed data on secondary emission are sufficient to permit drawing a qualitative picture of the

mechanism of the release of electrons. A brief summary of the salient points of these data is given in the next few paragraphs.

The secondary-emission ratio of most clean metal surfaces is quite low, that is, in the neighborhood of one or two. A few exceptions may exist for which the ratio is as high as four or five, but such values have been observed only for metals which are extremely difficult to free from contamination, even with the best technique known. Calcium, beryllium, barium and the alkali metals are examples of the apparent exceptions. Recently some observations have been reported* on caesium purified by careful multiple distillation, which indicate a very low ratio, whereas heretofore it had been considered one of the best pure metal emitters. This lends validity to the generalization that pure metal surfaces have a low secondary-emission ratio.

On the other hand, a very small amount of surface contamination may raise the emission to a high value. For example, a layer of oxygen alone is sufficient to double or triple the emission from such metals as magnesium, aluminum, or beryllium. Complex surfaces, similar to those used as photoemitters, consisting of caesium oxide and caesium, or rubidium oxide and rubidium, are among those exhibiting the highest ratio, being capable of emitting more than ten secondary electrons per primary electron. Very thin layers of alkali halides on certain metals also have very high ratios, even higher than the complex caesium surface mentioned above, although they are much less stable than the latter.

The velocity of the bombarding electrons is important in determining the secondary-emission ratio of a given surface. In general the ratio is small at low voltages, increasing with an increase in voltage to a maximum at 300 to 600 volts for the primary beam, after which it decreases slowly with a further increase in bombarding voltage. Figs. 1.15 and 1.16 show the variation of ratio with bombarding voltage. Fig. 1.15 illustrates the effect for a few pure metal surfaces†, typical complex surfaces being shown in Fig. 1.16.

Apart from the velocity of the bombarding electrons and the nature of the surface, the secondary-emission ratio depends upon the angle of incidence of the primary beam. At normal incidence the ratio is a minimum and increases as the angle tends towards grazing incidence. The rate of increase of yield with angle of incidence is greater for high-voltage primary electrons than for low. The form of the dependence upon bombarding angle can be seen from the measurements of H. O. Müller‡, shown in Fig. 1.17.

* See Bruining, references 16 and 19.

† See Kollath, reference 11.

‡ See Müller, reference 18.

Observations on the velocity of emission of secondary electrons reveal that the electrons may be conveniently classified in three groups. First there is a group of electrons which leave the surface with essentially the same velocity as that of the primary electrons. These electrons are often spoken of as reflected electrons. In general, this group constitutes but a small fraction of the total emission. Second, there is a group of low-

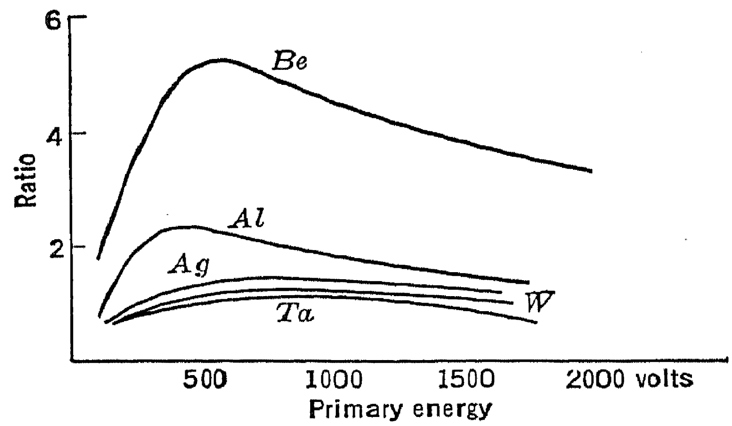


FIG. 1.15.—Secondary Emission Ratio for Various Metals (after Kollath).

velocity electrons. Finally, there is a background of electrons with velocities distributed more or less uniformly between the two. The low-velocity group is the primary concern of the present discussion because it makes up the bulk of the emission. An idealized velocity-distribution curve is given in Fig. 1.18, illustrating the three types of emission.

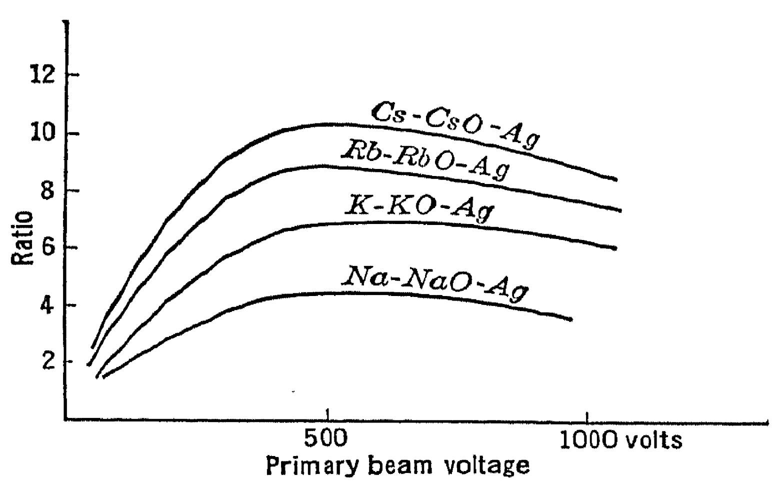


FIG. 1.16.—Secondary Emission Ratio of Complex Alkali Surfaces.

For the pure metals, where the secondary emission is relatively inefficient, the distribution rises sharply to a maximum at a relatively low velocity (usually under 5 volts) and then decreases slowly until it merges with the background at 25 to 50 volts. The velocity distribution of the low-velocity group for molybdenum is shown in Fig. 1.19. The velocity distribution of the more efficient emitters is much narrower than that

of the pure metals. The electron-optical properties, equilibrium potential determinations, and retardation measurements indicate that more than 85 per cent of the emitted electrons from a good caesiated-silver emitter (ratio 8 to 10) have velocities under 3 volts. Special emitter surfaces containing magnesium and magnesium oxide having a ratio of 6 or 7 have a distribution width about twice that of caesiated silver. The shape of the distribution curves resembles somewhat that of a Maxwellian energy curve but differs from it in detail.

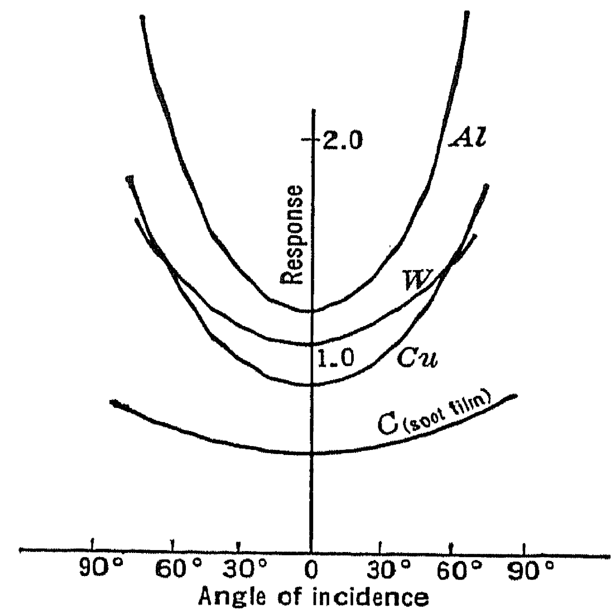


Fig. 1.17.—Variation of Secondary Emission Ratio with Angle (after Müller).

Accurate measurements of the velocity distribution are very difficult. Scattering effects, contact potentials, uncertainty as to the exact state and composition of the surface under examination, all contribute towards making the results inconclusive. For these reasons there is a paucity of reliable information on this phase of the secondary-emission phenomena. The information available on secondary emission is not sufficient to permit an unambiguous explanation of the phenomenon. Until further data on this form of emission are forthcoming, it is possible to formulate only tentative mechanisms which account for at least some of the features observed experimentally.

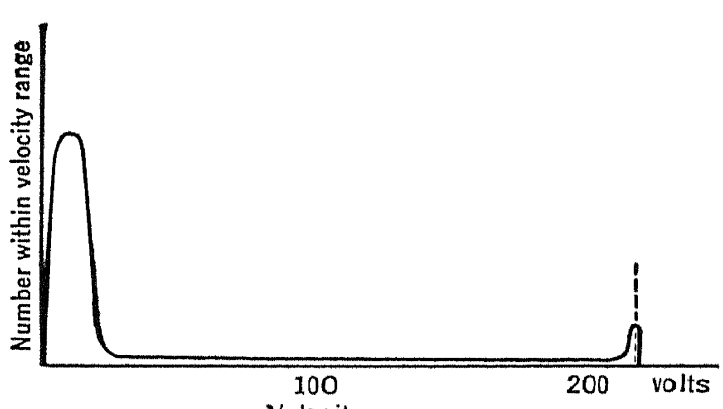


Fig. 1.18.—Idealized Velocity Distribution of Secondary Electrons.

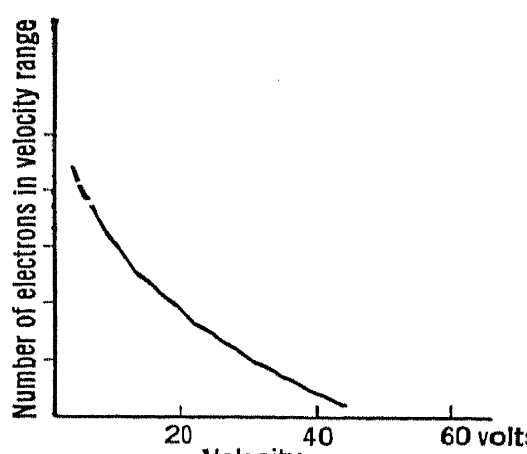


Fig. 1.19.—Velocity Distribution of Secondary Emission from Molybdenum.

Like the previously described forms of emission, secondary emission

requires that electrons in the metal be given a velocity which has a normally directed component of sufficient magnitude to permit their escape through the potential barrier. This energy is obtained through a collision or interaction between the incoming electron and electrons in the metal. Obviously, since secondary emission ratios of ten or more can be obtained, an incident electron must give its energy directly or indirectly to many internal electrons. Probably only a small fraction of the electrons which receive this energy actually find their way through the surface.

The assumption of a collision between an incident electron and a free electron in the metal cannot account for the latter acquiring velocity in a direction which would permit its escape. This is so, of course, because in an elastic collision between two electrons momentum must be conserved, just as it is for any other type of particle. The momentum vector of the electron struck must therefore make an angle of less than 90° with that representing the momentum of the incident electron before the collision. The experimentally observed initial velocities of the emitted electrons have components of motion opposite to that of the bombarding electrons and necessitate the assumption that a third element is involved in the collision. A number of mechanisms will suffice to account for motion in this direction—for example, multiple collisions, an interaction between the lattice and the incident or emitted electron, or a collision between the incident electrons and internal electrons which are bound to the lattice. H. Fröhlich in his interpretation of secondary emission in terms of quantum mechanics shows that there is a possible interaction between the atoms of the metal lattice and the conduction electrons which is sufficient to account for a yield in the neighborhood of unity. This theory explains much that is observed in the case of pure metal surfaces, but is scarcely applicable to compound surfaces.

The increased secondary-emission ratio observed when a metal surface is contaminated with foreign atoms may be due to an increased probability of interaction. This increase in interaction is possibly the result of a partial binding of electrons by the surface layer. Such a surface layer may also reduce the work function of the surface, although this is of less importance. There are many instances wherein the secondary-emission ratio is very greatly increased by a surface layer which definitely increases the work function. For example, a pure tungsten surface whose work function is 4.52 volts has a maximum ratio of about 1.5 or less. When it is contaminated with a layer of oxygen which raises the work function to 9.25 volts, the ratio is increased. Oxygen on aluminum, beryllium, or magnesium, etc., acts in the same way. The contamination of the surface with a material which definitely lowers the

work function also greatly increases the ratio. The effect in this case is probably twofold, being due both to an increased interaction between the electrons and to a reduction in work function.

Surfaces consisting of oxidized silver treated with caesium give a very high yield. The final step in the activation of such a surface is exposure to caesium. Therefore a layer of uncombined alkali is present, giving the characteristic low work function. If, after activation, the surface is exposed to oxygen, thus eliminating the free caesium and increasing the work function, the emission ratio drops by nearly 50 per cent. The lowering of the ratio in this instance is probably primarily due to the increase in work function rather than to any major change in the probability of interaction.

The relation between the probability of excitation and the velocity of the incident electron is rather complicated and not completely understood. Each excited electron removes some energy from the incident electron, so that if the latter has a given amount of energy it can excite only a limited number of metal electrons. Therefore the total number excited by a bombarding electron increases with bombarding voltage. However, the probability of excitation per unit length of path of the incident electron decreases as the voltage increases. At low voltages the incoming electron loses its energy very near the surface, and excited electrons which have velocities of the required magnitude and direction do not have to travel far in the material before escaping. In this velocity range, therefore, the yield increases with voltage. At high voltages the incident beam penetrates further into the emitter, and, although the total number of excited electrons produced is greater, most of them originate at depths too great to permit their escapes; hence above a certain optimum bombarding voltage the emission ratio decreases. If the high-velocity electrons enter the surface at a small angle the depth of origin of the excited electrons is less than for normal incidence, and consequently the secondary-emission ratio is greater.

De Boer, Bruining, and others have proposed a somewhat different mechanism, based on their extensive experimental work in the field of secondary emission. Some of the salient features of the explanation suggested by this school of thought are:*

1. When primary electrons fall on a metal surface they excite principally electrons which are bound to the lattice atoms rather than conduction electrons.

2. Although large numbers may be excited, relatively few escape, because of the high absorption in metals. This absorption is due to the

* See De Boer and Bruining, references 3, 16, 17, and 19.

existence of the partially filled permitted energy band which characterizes a metal.

3. Good emitting surfaces are thin insulating films, which, having no unfilled permitted energy band, offer little absorption to the excited electrons, so that a large number can leave the material.

4. To be a stable secondary electron source, the insulating layer must have means for replacing the electrons which have been emitted.

5. The work function plays a relatively unimportant role in this form of emission because of the high initial velocities.

Secondary emission can be utilized in electronic devices to intensify an electron current. Its value for this purpose depends to a large extent upon its noise properties. Therefore this phase of the phenomenon warrants some consideration in the present discussion. Like the shot noise from a thermionic or photoelectric emitter, the noise from secondary emission is a statistical effect. If certain assumptions are made concerning the way in which the electrons are emitted, the magnitude of the fluctuation current can be calculated.

In striking the secondary emissive surface, each primary electron causes the release, on the average, of R electrons, where R is the secondary-emission ratio. However, this does not mean that each individual incident electron causes the emission of exactly R secondary electrons, although this possibility cannot be excluded a priori from consideration. In general there will be a certain probability $p(z)$ that a given primary electron will produce z secondary electrons. On this basis the probability that n incident electrons produce N secondaries will be:

$$P(N,n) = \sum_{\Sigma z_i = N} \prod_{i=1}^{i=n} p(z_i). \tag{1.17}$$

The primary electrons, whatever their source, will be subject to some kind of fluctuation, and therefore the number, n, arriving in a unit of time must be put on a probability basis. Let $P(n)$ be the probability that exactly n electrons arrive during a given unit of time, and let $\bar{n}$ be the average number per unit time. The probability that N electrons will leave the surface then becomes:

$$P(N) = \int_0^\infty P(n)\, P(N,n)\, dn. \tag{1.18}$$

Finally the mean square deviation from the expected number of secondary electrons $R\bar{n}$ will be:

$$\overline{\Delta N^2} = \sum_{N=0}^{\infty} P(N)\, (N - R\bar{n})^2. \tag{1.19}$$

This mean square deviation in the number of electrons emitted per unit time can, of course, be related to a mean square fluctuation current over any desired frequency band by

$$\overline{i^2} = 2e^2\overline{\Delta N^2}\,(f_2 - f_1),$$

as was shown in an earlier section.

Before this expression representing the noise effective over a given frequency band can be evaluated, the various probabilities involved in its derivation must be determined. The probability $P(n)$ for the incident beam depends upon the electron source. If it is a photoelectric cathode or an emitter of similar nature, the electron arrival will be purely random; in other words, the probability $P(n)$ can be expressed by the normal law:

$$P(n) = \frac{1}{\sqrt{2\pi\bar{n}}}\epsilon^{-\frac{(n-\bar{n})^2}{2\bar{n}}}. \tag{1.20}$$

Assigning values to the probabilities $p(z)$ involves assumptions which imply various mechanisms of secondary emission.

One assumption, and perhaps the simplest, is that:

$$\begin{aligned} p(z) &= 1, \qquad & z &= R, \\ p(z) &= 0, \qquad & z &\neq R; \end{aligned}$$

or, in other words, that every primary releases just R secondaries. From this it follows that:

$$\begin{aligned} P(N,n) &= 1, \qquad & N &= Rn, \\ P(N,n) &= 0, \qquad & N &\neq Rn. \end{aligned}$$

Eq. 1.18 can now be written:

$$P(N) = \frac{1}{\sqrt{2\pi\bar{n}R^2}}\epsilon^{-\frac{(N-R\bar{n})^2}{2R^2\bar{n}}}. \tag{1.21}$$

The mean square deviation, according to the ordinary principles of probability, is:

$$\overline{\Delta N^2} = R^2\bar{n}, \tag{1.22}$$

or, in terms of current fluctuation,

$$\overline{i_n{}^2} = 2eR^2i_0\,(f_2 - f_1), \tag{1.23}$$

where i_0 is the primary current.

This result can be obtained without resorting to the above calculation by merely considering that the expected noise is that due to the random emission of particles of charge eR, producing a current $i = i_0R$. The cur-

rent fluctuations predicted by Eq. 1.23, however, do not fit the data derived from noise measurements, indicating that the assumptions used in its derivation are not valid.

A more reasonable assumption is that secondary emission itself is a random phenomenon. If the primary beam entering the target material excites a large number of electrons, only a few of which have velocities such that they are able to escape, it is not unreasonable to postulate the following two conditions:

1. The probability P of emission of any one of the m electrons excited by the beam is small compared with unity.

2. The emission of a secondary electron does not appreciably alter the probability of emission of the remaining excited electrons.

Assuming that these two conditions are fulfilled, the binomial theorem of statistics expresses the probability $p(z)$ that z electrons will be produced by a given incident electron. Therefore:

$$p(z) = \frac{m!}{z!\,(m-z)!}\, p^z(1-p)^{m-z}, \tag{1.24}$$

where the expectation mp is, of course, equal to R.

The probability of obtaining N secondary electrons from n primary electrons, which can be found with the aid of Eqs. 1.24 and 1.17, is:

$$P(N,n) = \frac{(nm)!}{(nm-N)!\,N!}\, p^N(1-p)^{nm-N}. \tag{1.25}$$

Since both mn and N are large and p is small compared with unity, Eq. 1.25 can, by means of Stirling's formula, be reduced to:

$$P(N,n) = \frac{1}{\sqrt{2\pi nR}}\,\epsilon^{-\frac{(N-nR)^2}{2nR}}. \tag{1.25a}$$

Again by making use of a perfectly random primary electron beam (i.e., Eq. 1.20 and Eq. 1.25a in Eq. 1.18), the probability of obtaining N secondary electrons from an average of $\bar{n}$ incident electrons will be:

$$P(N) = \int_0^\infty \frac{1}{\sqrt{2\pi\bar{n}}}\,\epsilon^{-\frac{(n-\bar{n})^2}{2\bar{n}}}\, \frac{1}{\sqrt{2\pi nR}}\,\epsilon^{-\frac{(N-nR)^2}{2nR}}\,dn$$

$$= \frac{1}{\sqrt{2\pi\bar{n}(R+R^2)}}\,\epsilon^{-\frac{(N-\bar{n}R)^2}{2\bar{n}(R+R^2)}}. \tag{1.26}$$

Therefore the mean square deviation is:

$$\overline{\Delta N^2} = \bar{n}\,(R^2+R), \tag{1.27}$$

and, by the same reasoning as before, the mean square current fluctuation over a frequency band $(f_2 - f_1)$ when the primary current i_0 strikes the target is expressed by:

$$\overline{i_n^2} = 2e(R^2 + R)\, i_0(f_2 - f_1). \tag{1.28}$$

This relation is in close agreement with results obtained from efficient emitting surfaces operated at voltages below that required to give maximum yield.

Comparing Eqs. 1.11 and 1.28, it will be seen that 1.28 consists of two parts, one giving the multiplied noise in the primary current, the other, which has the form

$$\overline{i_{n'}^2} = 2eRi_0(f_2 - f_1), \tag{1.29}$$

expressing the noise generated by secondary emission itself.

For poor emitters, or good emitters operated at high voltages, this latter portion of the noise is increased, so that in the general case it should be written as:

$$\overline{i_{n'}^2} = 2ekRi_0(f_2 - f_1), \tag{1.29a}$$

where k has a value close to unity for good emitters operated under conditions corresponding to those found in practice where secondary emission is used for current intensification.*

The explanation† brought forward to account for the departure of k from unity is that under conditions of poor emission an appreciable number of primaries strike the target and produce no true secondary electrons, because they are either absorbed by the emitter, or they are reflected from it. Owing to this, the effective primary beam is decreased, and the ratio for those electrons which do cause emission is greater than the observed secondary-emission ratio R. If this is the correct explanation, the factor k is not independent of R. Taking P_L as the probability of an electron being lost, k has the form:

$$k = \frac{P_L\, R}{1 - P_L} + 1.$$

For engineering noise calculations in cases of practical secondary-emission intensification, k may be taken as unity.

Secondary-emission multipliers in general consist of cascaded stages, each intensifying the current from a preceding stage. Diagrammatically a multiplier phototube can be represented as shown in Fig. 1.20. Electrons originate as photoelectrons from the cathode of the tube. These are

* The factor k corresponds to p used in Zworykin, Morton, and Malter, reference 14, and to bm in Shockley and Pierce, reference 22.

† See Kurrelmeyer and Hayner, reference 21.

directed, by a suitable electron-optical system, onto the first target, where they produce secondary electrons. This secondary emission is directed to the second target, and in turn generates more secondaries.

The multiplier just described, if the initial photocurrent is i_0, will have a current output of:

$$I_t = R^n i_0,$$

where R is the secondary-emission ratio of each stage and n the number of stages. The noise contributed to the output can be calculated for each

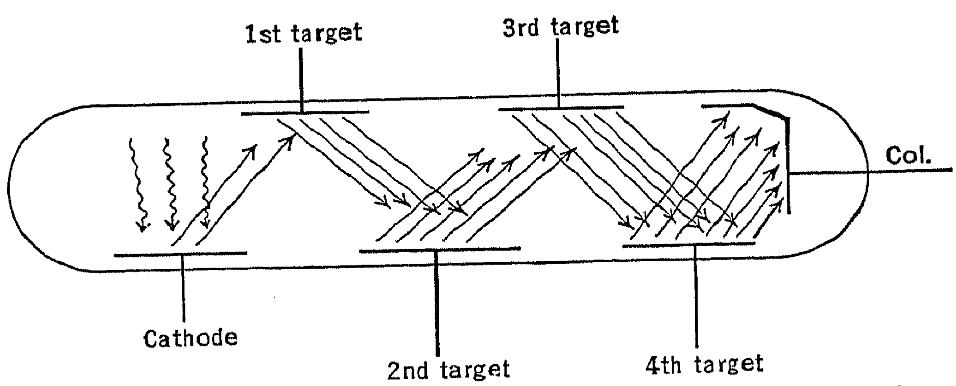


FIG. 1.20.—Secondary Emission Multiplier (Schematic).

successive stage, as shown in the following table, and the total noise found by summation.

Noise from photocathode $i_{n0}^2 = R^{2n}\, 2eFi_0 = R^n\, 2eFI$
Noise from first target $i_{n1}^2 = R^{2(n-1)} 2eFRi_0 = R^{n-1}\, 2eFI$
Noise from second target $i_{n2}^2 = R^{2(n-2)} 2eFR^2i_0 = R^{n-2}\, 2eFI$

Noise from nth target $i_{nn}^2 = 2\, e\, F\, R^n\, i_0 = 2eFI$

The total noise is therefore:

$$i_n^2 = (R^n + R^{n-1} + R^{n-2} \cdot \cdot \cdot + 1)\, 2eFI$$
$$= \frac{R^{n+1} - 1}{R - 1}\, 2eFI \cong \frac{R^{n+1}}{R - 1}\, 2eFI.$$

In order to compare the noise given by this equation with the actual noise measured from several multipliers, the following measurements were made:

No. of Stages (n)	Overall Gain R^n	Observed $i_n^2/2eFI$	Calculated $i_n^2/2eFI$
3	60	77	80
3	28	40	41
3	6.8	12.1	13.3
2	29.5	26.2	36.0
1	6.0	7.2	7.0

The measurement is sufficiently good to indicate that for practical noise calculations, such as are required in television or radio technique, these equations are adequate. These data are not sufficiently accurate, however, for an exact determination of k. On the basis of the measurements given, k is found to vary over a range of values which lie within $\pm$ 20 per cent of unity.

Since secondary emission is rapidly increasing in importance as a means for intensifying small electron currents, a discussion of the method of activation for caesiated silver emitting surfaces is not out of place. The preparation follows closely that used in producing a caesiated silver photoemitter. The target to be processed may be made of silver, or of any other metal plated with a heavy layer of silver. The silver should be relatively free from metallic contamination, such as alloyed copper. Before the activation is started the surface of the target must be thoroughly cleaned, both before and during the exhaust of the tube containing the target. The surface is then oxidized by the same procedure employed for the corresponding photoelectric emitter. After the tube has been pumped free from oxygen, caesium is admitted. The amount of caesium needed to obtain maximum yield is somewhat less than that required to produce an optimum photoemitter, the quantity required being dependent upon the degree of oxidation. The tube is next baked at 200°C to permit a reaction between the silver oxide of the target and the caesium. The activated surface thus formed should have a brown color.

The response, unlike that for a caesiated silver photoemitter, is not increased by silver sensitization, that is, by evaporating a thin layer of silver on the finished surface. Such a procedure reduces the secondary emission ratio to less than half its optimum value.

The similarity between the photoelectric and secondary-emission activation of silver is of considerable advantage in the processing of the photoelectric secondary-emission multiplier. Because of it, it is possible, by oxidizing the silver of the targets slightly more than that of the photocathode, to activate the two types of surfaces simultaneously.

Other secondary-emission surfaces, giving a high yield, are also of importance. Caesium-activated beryllium-beryllium oxide, and rubidium-activated silver-silver oxide, are both excellent emitters. Various alloys can be formed which, when activated, will give relatively high emission ratios, even at temperatures above 500°C. These alloys are of particular value for multipliers designed to handle large amounts of power.

Insulators and materials of very low conductivity also emit secondary electrons. The range of secondary-emission ratio is about the same as that for conducting materials, extending from 10 or 11 for caesium-treated mica down to less than unity for certain of the fluffy oxides. The

ratio varies with the velocity of the incident electrons in much the same way as it does for conducting emitters. When an insulator is bombarded with electrons its surface must assume a potential such that the net current to or from it is zero, since electrons will not flow through the material to compensate for any change in the number at the surface. Two potential equilibria fulfill this condition: the first is for the surface to become so negative that the primary electrons cannot reach it; the other is for it to assume a potential which is slightly positive with respect to the element collecting the emitted electrons so that the secondary-emission current is not saturated, and the current which leaves the surface is just equal to the primary current. The latter is possible only when the saturated secondary-emission ratio is greater than unity.

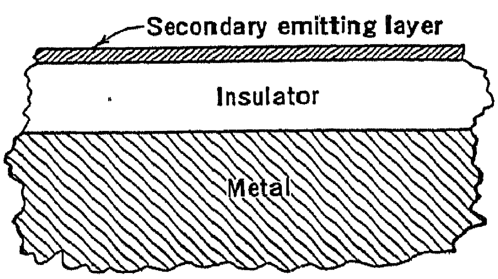


FIG. 1.21.—Thin Film Field Emitter.

Because current does not flow through a target made of insulating material, measurements of the secondary-emission ratio must be made by ballistic methods. In making these measurements advantage can be taken of the capacitance between the surface of the target and a conducting surface backing it up.

Directly or indirectly, the secondary-emission properties of dielectrics are of considerable practical interest. The secondary emission from the glass walls of electronic devices often plays a role in their operation. The phosphors used in cathode-ray oscilloscopes and television receiving tubes to transform the energy of primary electrons into light are non-conductors. Their secondary-emission properties are therefore important in determining their performance, as will be discussed in Chapter 2.

A phenomenon closely related to secondary emission is thin film field emission.* The type of surface exhibiting this effect can be represented diagrammatically as in Fig. 1.21. Two distinct layers make up the complex surface of this emitter. The outermost layer, which has a very high transverse resistance, is activated so as to have a secondary-emission ratio greater than unity. Under it is a thin insulating film. When the outer surface is bombarded it tends to assume a potential approximately equal to that of the collector, as was explained in the discussion of the secondary emission from insulators. When the collector is at a positive potential with respect to the target, a very high gradient will exist between the metal surface and the outer layer. By using an extremely thin insulating film, the field developed can be made sufficiently intense to draw electrons from the metal and cause them to pass through the di-

* See Malter, reference 23.

electric into the space beyond. The ratio of emitted current to primary current for this type of surface may be several thousand. However, the emitted current responds very slowly to changes in primary current owing to the storage of charge on the surface. This limits the utility of the phenomenon.

A very similar phenomenon was observed somewhat earlier by Güntherschulze* in gas discharges when the cathode was coated with an insulating powder.

REFERENCES

1. F. K. Richtmyer, "Introduction to Modern Physics," McGraw-Hill, New York, 1934.
2. A. L. Reimann, "Thermionic Emission," John Wiley and Sons, New York, 1934.
3. J. H. de Boer, "Electron Emission and Adsorption Phenomena" (Cambridge University Press), Macmillan, New York, 1935.
4. T. J. Jones, "Thermionic Emission," Methuen and Co., London, 1936.
5. V. K. Zworykin and E. D. Wilson, "Photocells and Their Application," John Wiley and Sons, New York, 1934.
6. A. L. Hughes and L. A. DuBridge, "Photoelectric Phenomena," McGraw-Hill, New York, 1932.
7. L. R. Koller, "Physics of Electron Tubes," McGraw-Hill, New York, 1937.
8. Saul Dushman, "Thermionic Emission," *Rev. Modern Phys.*, Vol. 2, pp. 381–476, October, 1930.
9. W. Schottky, "Spontaneous Current Fluctuation in Emitted Electron Streams," *Ann. Physik*, Vol. 57, pp. 541–567, Dec. 20, 1918.
10. T. C. Fry, "Theory of the Schrot-effekt," *J. Franklin Inst.*, Vol. 199, pp. 203–220, February, 1925.
11. R. Kollath, "The Secondary Emission of Solid Bodies," *Physik. Z.*, Vol. 38, pp. 202–224, April, 1937.
12. R. Warnecke, "The Emission of Secondary Electrons from Metallic Surfaces," *Onde électrique*, Vol. 16, pp. 509–540, September, 1937.
13. R. Warnecke, "Secondary Emission of Pure Metals," *J. phys. radium*, Vol. 7, pp. 270–280, June, 1936.
14. V. K. Zworykin, G. A. Morton, and L. Malter, "The Secondary Emission Multiplier," *Proc. I. R. E.*, Vol. 24, pp. 351–375, March, 1936.
15. H. Fröhlich, "Theory of the Secondary Emission of Metals," *Ann. Physik*, Vol. 13, pp. 229–248, April, 1932.
16. H. Bruining and J. H. de Boer, "Secondary Electron Emission," Part I, "Metals," *Physica*, Vol. 5, pp. 17–30, 1938.
    H. Bruining, Part II, "Absorption of Secondary Electrons," *Physica*, Vol. 5, pp. 901–912, December, 1938; Part III, "Slow Electron Bombardment," *Physica*, Vol. 5, pp. 913–917, December, 1938.
17. H. Bruining and J. H. de Boer, "Secondary Electron Emission of Metals with Low Work Function," *Physica*, Vol. 4, pp. 473–477, June, 1937.
18. H. O. Müller, "Dependence of Secondary Emission from Metals upon Bombarding Angle," *Z. Physik*, Vol. 104, pp. 475–486, 1937.

* See Güntherschulze, reference 24.

19. H. Bruining, "Secondary Emission from Solids" (Dutch), Thesis, Leiden, 1938; N. V. Bock, Leiden, 1938.
20. M. Ziegler, "Fluctuation Effects in Secondary Emission," *Physica*, Vol. 3, pp. 1–11, January, 1936; pp. 307–316, May, 1936.
21. B. Kurrelmeyer and L. J. Hayner, "Shot Effect of Secondary Electrons from Nickel and Beryllium," *Phys. Rev.*, Vol. 52, pp. 952–958, Nov. 1, 1937.
22. W. Shockley and J. R. Pierce, "A Theory of Noise for Electron Multipliers," *Proc. I. R. E.*, Vol. 26, pp. 321–332, March, 1938.
23. L. Malter, "Thin Film Field Emission," *Phys. Rev.*, Vol. 50, pp. 48–58, July 1936.
24. A. Güntherschulze, "Electron Velocity in Insulators at High Field Strength and Its Relation to the Theory of Electric Penetration," *Z. Physik*, Vol. 86, pp. 778–786, Dec. 7, 1933.

CHAPTER 2

FLUORESCENT MATERIALS

Strictly speaking, there is no doubt that the title of this chapter is a misnomer. More exactly, the phenomenon to be discussed is that of cathodoluminescence, and the materials which manifest this effect are known by the general name of phosphors. However, the use of the term "fluorescent materials," instead of the more exact name, is so widespread that the latter seems almost an affectation.

2.1. Luminescence. The conversion of energy invisible to the unaided eye into visible light is a very common phenomenon. When this conversion takes place at a temperature less than that required for incandescence, the emission of light is known as luminescence. Luminescence is commonly subdivided into groups classified according to the means of excitation. Table 2.1 lists the more important groups, together with the

TABLE 2.1

Class	Excitation
Photoluminescence	Radiant energy; visible part of spectrum; x-rays; ultra-violet.
Cathodoluminescence	Electron bombardment.
Radioluminescence	Radiation from radioactive sources.
Triboluminescence	Disruption of crystals, as by grinding, abrasion, etc.
Bioluminescence	Biochemical reactions.
Chemiluminescence	Chemical reactions.
Thermoluminescence	Temperature-excited light in excess of black-body radiation.

means of excitation. Furthermore, luminescence—irrespective of the means of excitation—is divided into fluorescence and phosphorescence. Expressed inexactly, fluorescence is luminescence which ceases almost immediately upon removal of the excitation, whereas in phosphorescence the luminescence persists. It is difficult to give a more rigorous definition of these terms since, as yet, there is no generally accepted criterion to distinguish between them. The most commonly used time limit for the duration of fluorescence after cessation of excitation is 10^{-8} second (i.e., the average time required for an excited atom to return to its normal state).

Cathodoluminescence, as can be seen from Table 2.1, is excited by bombarding the material with electrons. There are a vast number of substances which have the property of luminescing when thus bombarded.

In fact, nearly all inorganic compounds which are non-metallic crystals exhibit this effect to some extent, together with a great many organic compounds, non-metallic elements, and glasses.* However, for most materials the efficiency of the production of light is extremely low, and, therefore, they are of no interest in the present consideration, which will be restricted to the relatively few materials, natural or synthetic, which are efficient phosphors.

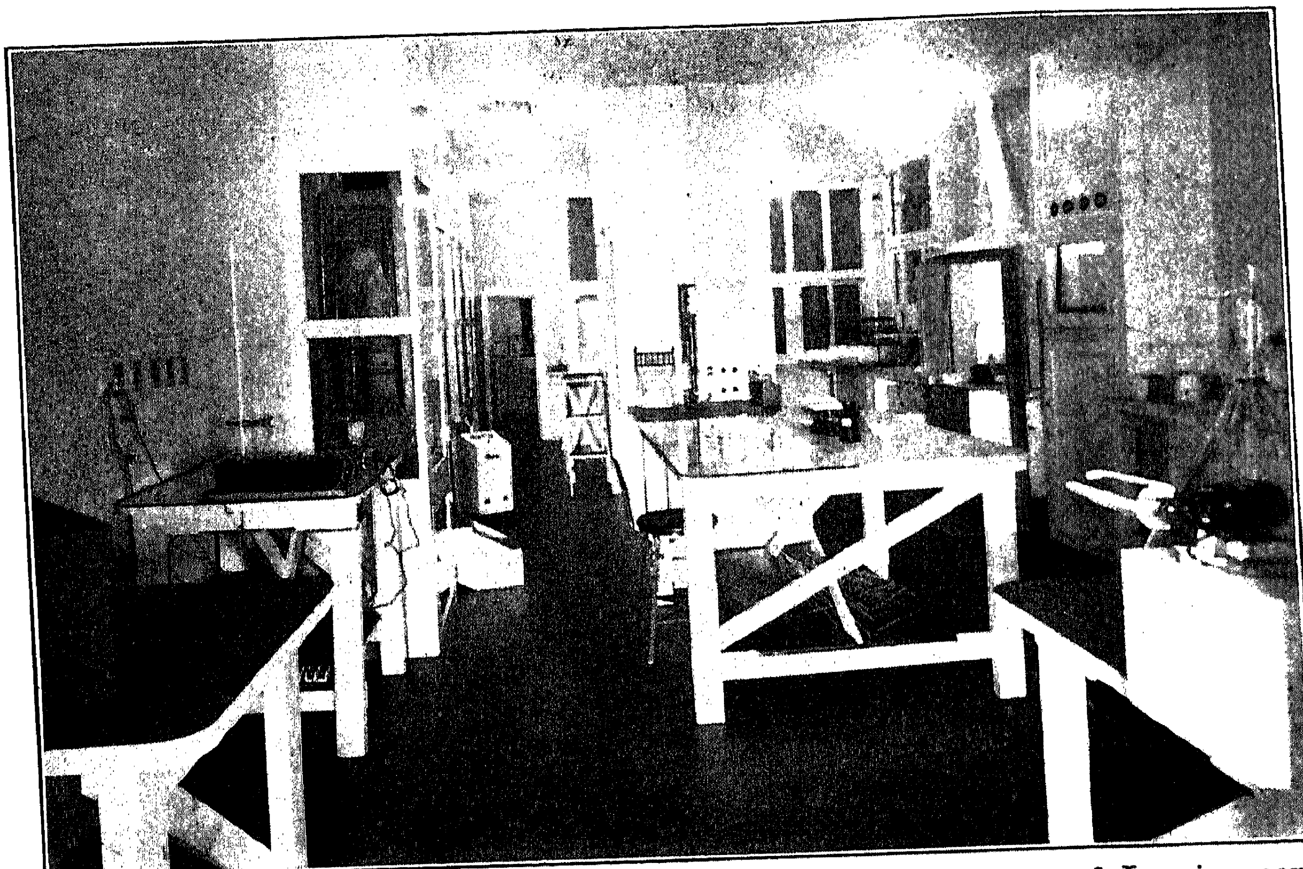


Fig. 2.1.—A Modern Research Laboratory for the Investigation of Luminescent Materials (RCA Manufacturing Company).

2.2. Requirements of a Phosphor. In television work a phosphor is used in the form of a screen which serves to convert the energy in a scanning electron beam into the light which makes up the reproduced picture.

One of the prime requisites of the screen material is that it be an efficient converter of energy. In addition to this requirement it must have the following properties:

The ability to produce high, instantaneous, intrinsic brightness.

Stability and long life under electron bombardment.

Suitable electrical properties.

* Weak cathodoluminescence has been reported for certain metals by R. Tomaschek.

A low vapor pressure.
A color pleasing and restful to the eye.
A duration of phosphorescence not longer than a picture period.

The above seven items are essential, but certain other properties are highly desirable. For example, the luminous output should be approximately linear with current over a wide range, the material should be easy to apply, and its preparation should be commercially practical.

Until quite recently very little was known about the synthesis of phosphors. Early investigators in the field of cathode-ray television were forced to rely largely on naturally occurring materials such as willemite and zinc blende. However, it is now possible to synthesize phosphors which are far superior to the best naturally occurring material in every respect. Therefore synthetic substances will be the subject of most of what is to follow.

2.3. The Nature of Inorganic Phosphors. A few solids exhibit strong luminescence in their pure state. This statement must, of course, be qualified because there is no measure of absolute purity; however, the luminescence continues unchanged in these solids even when they are purified to such an extent that the spectrograph fails to reveal the presence of foreign material. Barium platinocyanide and zinc sulphide, for example, are such substances.

Most phosphors depend for their luminescence upon the presence of an "impurity," or activator. The activator influences not only the efficiency of the phosphor but also the color of its luminescence and the duration of phosphorescence. Therefore, by controlling the quantity and nature of the activator during the synthesis of the material, the desired properties can be obtained. The flexibility of this second class of phosphors has resulted in its wide use for practical television purposes.

In general, the second type of phosphor is made with three components. These are the base material, a flux, and finally the activator.

The base should be a colorless, electronic semi-conductor and must be crystalline in nature. Examples of such materials are ZnO, CdO, MgO, Al_2O_3, TiO_2, CaO, SiO_2, CdS, and ZnS, all of which are good phosphor bases. These materials all exhibit semi-conductivity due to the motion of electrons through the crystal. Substances which have a sharply defined cation-to-anion ratio, such as the alkali halides, are very poor cathodoluminescent bases. Likewise, semi-conductors which do not depend upon the motion of electrons for their conduction, such as Cu_2O, FeO, NiO, and $CuCl$, do not form efficient bases.

Summarizing the qualifications of good base materials, they must be nearly colorless compounds which have their cations combined in their

highest (and, preferably, only) valence states, and which tend to have an excess of cations. In addition, the material must be sufficiently stable not to disintegrate under the electron beam.

The second component, the flux, is not necessary in the preparation of all types of phosphors. This material, usually an alkali or alkaline-earth halide, plays a role only during the crystallizing process. The flux aids in the crystallization of the substance and is removed subsequent to the final heating after it has completed its pseudocatalytic action.

A small quantity of certain metals, say 1 atom to every 100,000 atoms of the base, added to a suitable base will tremendously increase the luminous efficiency of the base material. Such a metal is known as an activator. The theory of the activator is understood qualitatively only, but there is available a great deal of empirical data about the action of various elements as activators. The best activators are metallic elements having a multiplicity of valences. For example, Cu, Ag, Bi, Cr, and Mn are all excellent activators, although they do not, of course, behave uniformly toward all bases. The presence of more than one activator in a base usually results in a less efficient phosphor than if the best activator is used alone, but sometimes desirable color changes can be produced by the addition of a second element. When one activator is used the quantity present is not critical, and the curve of efficiency as a function of the amount of activator usually exhibits a very broad maximum.

It is interesting to note that certain elements, known as "killers," or poisons, when present even in an amount small compared with that of the activator, result in a very great reduction in the luminous efficiency of the phosphor. Fe, Co, Ni, and Pb all behave in this way. Their action is to suppress the processes of phosphorescence and thereby reduce the total emitted energy. Minute quantities of these elements are often used in the preparation of fluorescent screens where a short time lag is desired.

The exact method of synthesizing a phosphor depends upon the particular material; however, there is a basic procedure common to most phosphors.

The base material, together with the flux, when used, are ground together in a clean, inert crucible. To this is added the activator required to produce the desired color—generally in the form of an aqueous solution of the salt of the metal used as activator. After drying and regrinding, the mass is heated and allowed to crystallize. The phosphor may then be put through a final grinding process to comminute the crystal particles to the size required for application to the screen of the viewing tube.

2.4. General Properties of Inorganic Phosphors. Nearly every step in the preparation and subsequent handling of a phosphor has a marked

influence on its properties. This includes not only the kind and quantity of activator but also the crystallization procedure, the grinding, the composition of the base, etc. Some of these effects are illustrated in the following examples.

The spectral output of zinc orthosilicate under different conditions of preparation is shown in Fig. 2.2. Curve (1) is the luminescence of the pure base. The addition of a small amount of manganese results in curve (2*a*) if the material is cooled slowly. More of the manganese activator increases the luminous output without change in spectral characteristics. Eventually an optimum is reached represented by curve (2*c*), and further addition of manganese results in a gradual decrease in efficiency. If the

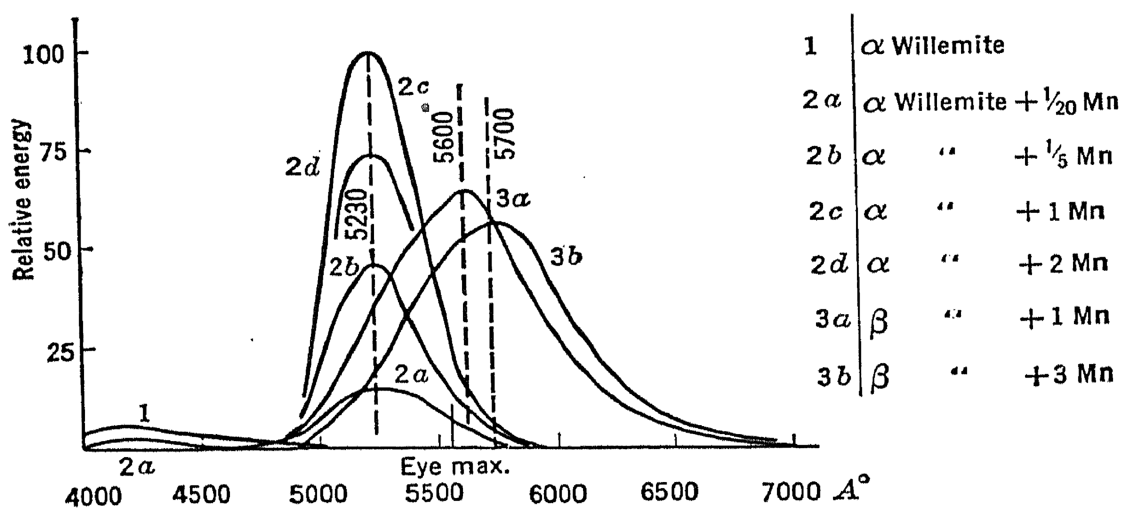


FIG. 2.2.—The Effect of Preparation and Activator on the Spectral Characteristics of Zinc Orthosilicate.

material, instead of being cooled slowly, is quenched rapidly from a melt, quite different results are obtained. An amount of manganese activator equal to that which previously resulted in the material giving curve (2*c*) now leads to a yellow phosphor shown in curve (3*a*). As the amount of activator is increased the color of the luminescence shifts to longer wavelengths. Curve (3*b*) is that of a material having three times the amount of manganese used to obtain curve (3*a*).

Another interesting series is that of a zinc cadmium sulphide base with a silver activator, the ratio of zinc to cadmium being varied. Fig. 2.3 shows a family of curves illustrating the behavior of this series. Pure zinc sulphide has the blue luminescence shown by curve (a_0). The addition of silver activator increases the energy efficiency and shifts the spectral output further into the blue [curve (a)]. As cadmium replaces zinc in the base the color response shifts toward the red. With increasing cadmium the output goes through a maximum and decreases until, for silver-activated cadmium sulphide alone, a rather weak, deep red lumi-

nescence is obtained. This is illustrated by the sequence of curves (*b*), (*c*), (*d*).

As in the two examples cited, a similar series of complicated phenomena is observed for nearly every phosphor group.

In addition to the crystal structure, the state of strain is an important factor in determining the luminous efficiency. For example, a freshly prepared phosphor which has high efficiency can be completely ruined if it is violently ground. Even a quite gentle grinding will reduce the efficiency to some extent because it will leave a strained surface layer over each grain.

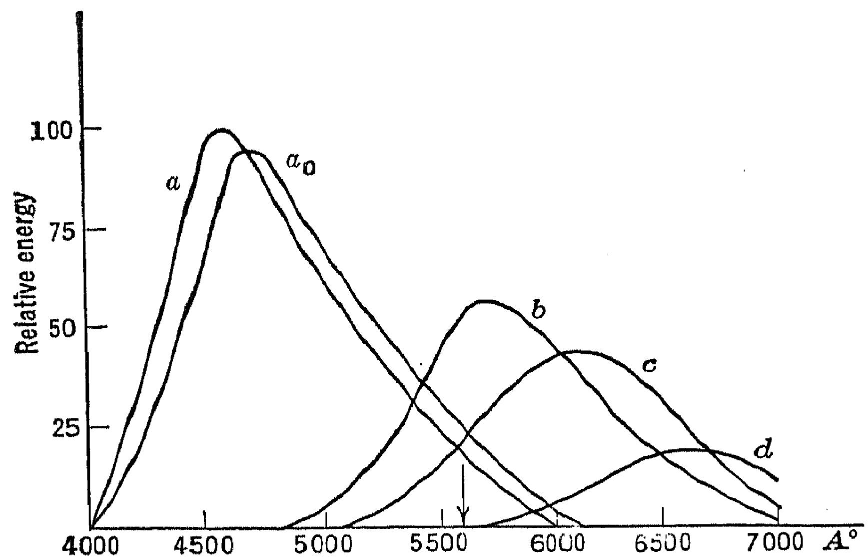


FIG. 2.3.—The Variation of Spectral Characteristics of Zinc Cadmium Sulphide with Zn : Cd Ratio.

The surface brightness of most phosphors is proportional to the current density of the bombarding electrons for low values of current density. As the density is increased the material saturates and the brightness rises less rapidly. The form of the saturation curve and the current value at which saturation begins depend upon the kind of material used and upon its treatment.

The relation between light output and beam voltage is rather difficult to determine for reasons that will be made clear in a later section. Available measurements indicate that over the working range of the material the light output from a phosphor is proportional to some power of the voltage, the exponent lying between 1 and 2.

In actual television practice, the beam which serves as excitation is on a given area for a length of time equal to about 0.1 or 0.2 microsecond. The light emitted from the area continues for several hundredths of a second, as can be seen from the persistence curves shown in Figs. 2.10 and 2.14. The greater part of the light energy, therefore, is released after

the beam is no longer exciting the area in question. In other words, the phosphorescence of the material provides most of the light. The total light energy released, i.e., the light intensity integrated over the time it persists, is approximately proportional to the energy supplied to the phosphor, rather than to the instantaneous power in the cathode-ray beam. Thus, if a beam of 10 μA and 6000 volts excites a given area for 10^{-6} second, the light averaged over 2/100 second will be approximately the same as that produced by an excitation at 100 μA for 10^{-7} second. The accumulative effect should not be considered as a rigorous law, and it exists only over periods of time which are short compared to the decay time and for currents well below the saturation of the phosphor. As will be explained in a later chapter, the effect has been applied to the type of picture transmission involving velocity modulation.

A final consideration in this general survey is that of the effect of temperature on luminescence. The thermal state of the phosphor has a much greater effect on phosphorescence than on fluorescence. At very low temperatures, fluorescence decreases and, in some cases, may alter its color radically; this occurs only at temperatures around that of liquid air. Between $-80°$C and $300°$C there is very little change in such phosphors as zinc orthosilicate and zinc sulphide, but if the temperature is carried much above 300°C the fluorescence decreases rather rapidly. The effect of a decrease in temperature on phosphorescence is to increase its duration and lower its intensity. Calcite is an excellent illustration of this. At room temperature, this substance has a brilliant orange phosphorescence which continues several minutes. If a calcite crystal is excited by cathode-ray bombardment and then cooled, the brilliancy decreases until it finally disappears. If it is again warmed up to room temperature the phosphorescence reappears without further excitation. If it is raised to a high temperature, the brightness increases and the duration decreases. An excited calcite crystal may be kept for days at liquid-air temperature, and then made to phosphoresce merely by bringing it up to room temperature. This behavior indicates that a certain quantity of energy available for phosphorescence is stored in the crystal, and that its rate of release in the form of light is a function of temperature. Thus, at a high temperature it is emitted very rapidly, giving brilliant luminescence and rapid decay, while at a low temperature this energy is released slowly over a long period of time.

In the next section, which briefly discusses the mechanism advanced by quantum physics to explain the phenomenon of luminescence, it will become evident why this behavior of phosphorescence is to be expected.

2.5. The Theory of Luminescence. To account for the phenomenon of luminescence it is necessary to make use of the quantum-mechanical

aspect of crystalline semi-conductors. This aspect of metals was discussed in the preceding chapter, and it was there concluded that electrons were distributed in an energy band whose characteristics are given by Fermi-Dirac statistics. For the present purpose it is necessary to consider not only the energy distribution of the electrons themselves, but also the energies it is possible for an electron to have.

The atoms in any crystal are so arranged that they produce a potential field having a three-dimensional periodicity corresponding to the lattice structure. Electrons moving in this field have kinetic and potential energies which are determined by their quantum state and their position. If the possible energies that these electrons can have are calculated on the basis of quantum mechanics, taking into account the exclusion principle, it is found that they can occupy only certain energy bands, there being other bands in which it is impossible for electrons to

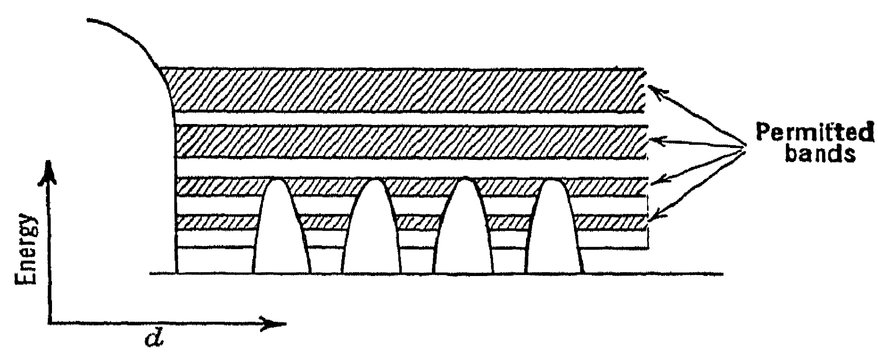


FIG. 2.4.—Energy Bands in a Crystal (Schematic).

exist. This is clarified by Fig. 2.4, which shows a single row of atoms, energy being indicated by the ordinate and position by the abscissa. The shaded areas in this figure are those in which electrons can exist; the unshaded regions are the so-called forbidden energy bands. The three-dimensional case of a real crystal, although much more complicated, is the same in principle.

In a metal the electrons only partially fill a permitted band, and therefore are capable of receiving very small increments of energy. This permits them to assume a drift motion under the influence of a field, which accounts for the observed conductivity. The system of permitted and forbidden bands in no way alters the picture of the electron behavior in metals given in Chapter 1.

When, however, the electrons just suffice to fill completely the permitted band they occupy, they are not capable of accepting an increment of energy which is less than that required to raise them to the next permitted band. The crystal will then be either an insulator or a semi-conductor, the distinction being one of degree rather than kind.

An insulator is a material having the next permitted band widely separated from the filled band; a semi-conductor, one in which the bands are close together.

Luminescent materials are, in general, semi-conductors and therefore have no partially filled permitted energy bands. Any incident electron cannot impart to the internal electrons an amount of energy less than that required to raise them to the next permitted band. This is illustrated in Fig. 2.5, which shows a simplified energy diagram with one filled band and one unoccupied permitted band, separated by an energy E. An arrow indicates the excitation of an electron, i.e., its transition from the lower to the upper level, caused by interaction with an incident electron. The excited electron moves through the crystal until it approaches a vacancy in the lower band. By giving up an amount of energy E, the electron occupies this vacancy. The energy released is in the form of electromagnetic radiation, whose frequency ν is given by the quantum relation $E = h\nu$, where h is Planck's constant. When this energy lies between $1\frac{1}{2}$ and 3 electron volts, the radiation falls into the visible portion of the spectrum.

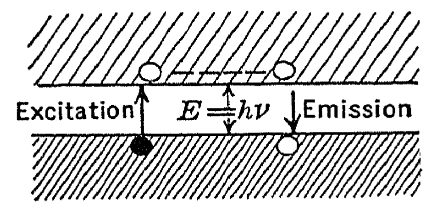


FIG. 2.5.—Energy Transitions Giving Rise to Luminescence.

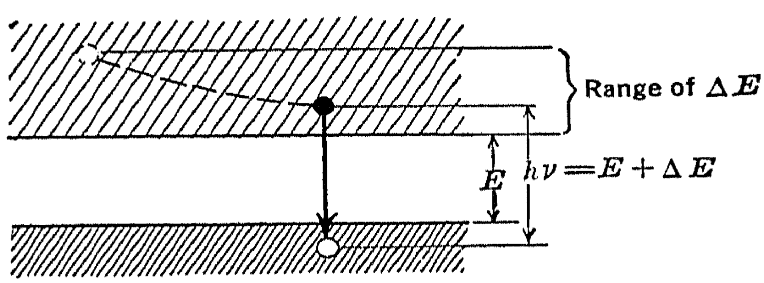


FIG. 2.6.—Energies Involved in the Emission of a Spectral Band of Luminescence.

In an actual crystal the energy change of the electrons in transition does not correspond exactly to the width of the forbidden band. This is because an electron may fall from a level somewhat higher than the bottom of the permitted band corresponding to the excited state and may terminate in a position below the top of the ground level. A transition of this type is illustrated in Fig. 2.6. The bombarding beam raises an electron well up into the unfilled level. As this electron moves through the crystal it loses energy to the lattice and drops down toward the bottom of the band. When it makes its transition corresponding to emission, its energy still differs from the lowest energy of the allowed level by a small increment ΔE. Accordingly, the radiation frequency is

$\nu = (E + \Delta E)/h$, where ΔE may have a range of values. The emission spectrum will therefore be in the form of a band.

The probability of the cycle considered so far is, in general, rather small. Even when the width of the transition region is such as to give rise to visible radiation, pure crystals are rarely good phosphors.

Foreign atoms in the crystal, either in lattice or interstitial positions, will have many permitted levels of their own, some of which may fall between the allowed levels of the base material. This is illustrated schematically in Fig. 2.7. The presence of the additional level, if unfilled, permits both excitation and emission to take place in two steps. If, on the other hand, the level due to the foreign or activator atom is occupied by an electron, this electron may be excited. The probability of the latter process is relatively high.

When an electron is excited and moves into a higher energy level it may be free to move through the crystal quite detached from the vacancy in the lower level created by its transition. Emission can in this case be the result of its returning to the ground level by occupying any vacancy which it happens to encounter. On the other hand, the excited electron may be constrained to remain in the vicinity of the vacancy it has created. Under these circumstances the emission of radiation is caused by the return of the electron to its original position. The first process is analogous to a bimolecular reaction; the second, to a monomolecular reaction. Both are observed in actual phosphors, being distinguishable by the form of the build-up and decay characteristics.

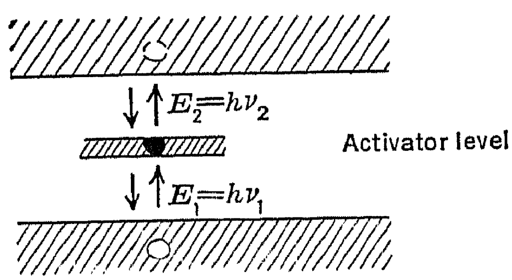


Fig. 2.7.—Electronic Transitions in the Presence of Impurities.

Phosphorescence and its dependence upon temperature require an interaction between the thermal motion of the lattice and the excited electrons. This interaction must be such that, as the thermal agitation becomes more violent, the probability of the electron losing energy and dropping to its original level becomes greater. It has been suggested that, where temperature-dependent phosphorescence exists, the transition cannot take place unless the active center has a certain position relative to the lattice, and that it is moved into or through this position as a result of thermal vibration. An actual deformation of the lattice as a result of the original excitation of the center may be involved in this mechanism.

It will be recognized that this discussion is highly speculative and contains many doubtful points; however, it is quite possible that the true explanation will be found to resemble the one given in its major aspects. The foregoing at least indicates the type of phenomenon involved. All attempts to give a quantitative picture of luminescence on

the basis of quantum mechanics have so far met with little success. As can be appreciated from the foregoing, the phenomenon is very complex, and any rigorous analysis must consider a three-dimensional potential distribution, the random location of the activator centers, the mosaic irregularities of a real crystal, and the effects of thermal motion.

2.6. The Theory of Luminescence (*continued*). Leaving the microphysical aspect of the problem, there are certain quantitative relations which can be deduced on a classical, statistical, basis,* which are helpful in a general survey of the subject.

Empirically, it has been found that for a number of the phosphors used in television tubes the following relation gives an accurate value for the light output over a wide range of voltages and currents:

$$L = A \cdot f(\rho) \cdot V^n, \tag{2.1}$$

where ρ is the current density and V the voltage. The exponent n has a value in the neighborhood of 2, though it differs somewhat for different materials, and even with the treatment of the material. The function $f(\rho)$ is found to be independent of voltage, at least over the range of bombarding voltages for which the power law is valid (that is, over the range from 500 to 10,000 volts). The form of this function is essentially linear for the lower values of ρ, but departs from linearity in the direction of decreasing efficiencies at high current densities.

Consider first the voltage relation. Under the assumption that the probability of excitation of a center by an electron with a velocity of V volts is $S(V)$, the light output from a unit area of thickness dx will be:

$$dL = Bf(\rho)S(V)dx. \tag{2.2}$$

It is known that (1) the electron density through any surface remains constant as the beam penetrates until the electrons are practically stopped, and (2) that the variation in velocity expressed in volts is $dV/dx = -K/V$. Therefore:

$$\begin{aligned} dL &= Bf(\rho)S(V)\left(\frac{dx}{dV}\right)dV \\ &= \frac{-B}{K}f(\rho)S(V)\ VdV. \end{aligned} \tag{2.3}$$

If the bombarding voltage is V_0, Eq. 2.3 leads to the following integral:

$$L = \frac{-B}{K}f(\rho)\int_{V_0}^{0} S(V)\ VdV. \tag{2.4}$$

* See S. T. Martin and L. B. Headrick, reference 10; and E. G. Ramberg and G A. Morton, reference 11.

Making the rather naive assumption that $S(V)$ is a constant, the light output becomes:

$$L = Af(\rho)V_0^2. \tag{2.5}$$

The assumptions made are far simpler than is justified, but the deduction is at least suggestive.

Experimental measurements seem to indicate that a relation exists between the light output as a function of current density and phosphorescent decay. Theoretically, there is reason to believe that such a relation may exist. Luminescence is the result of the return to the ground state of certain centers which have been excited by the bombarding electrons. On this basis, there is a certain probability that any given center which has been excited will not have returned to its ground state before the arrival of the next bombarding electron and thus will be incapable of further excitation. This probability increases with current density and with length of phosphorescent decay.

Phosphors for which the excited electron and the vacant position in the ground state remain associated exhibit a logarithmic phosphorescent decay. This form of decay curve is to be expected if it is assumed that the probability of emission from an active center during a small interval of time Δt is independent of time and the state of excitation of the remaining centers. The rate of recombination, under these circumstances, is:

$$\frac{dN}{dt} = -\alpha N, \tag{2.6}$$

where N is the number of excited centers and α a constant. If this expression is integrated, the number of active centers at any time t is found to be:

$$N = N_0 e^{-\alpha t}. \tag{2.7}$$

Since the intensity of the emitted light L is proportional to the rate at which centers recombine, the phosphorescence, as a function of time, is given by:

$$L = Ce^{-\alpha t}. \tag{2.8}$$

Under bombardment by an electron beam of current density ρ, an equilibrium is established between the rate of formation of excited centers and the rate of recombination. The number of excitations during the interval of time dt is proportional to the number of unexcited centers available and the current density of the incident beam. If, therefore, the

total number of available centers is M and the number of excited centers N_0 the condition of equilibrium can be expressed as:

$$\eta\rho(M - N_0)dt = \alpha N_0 dt; \tag{2.9}$$

therefore

$$N_0 = \frac{\eta\rho M}{\eta\rho + \alpha}. \tag{2.10}$$

From this the light output as a function of current density is:

$$L = D\frac{\rho}{1 + \beta\rho}. \tag{2.11}$$

When the mechanism of luminescence involves independent excited electrons and vacant energy states, the rate at which emission can occur is proportional to the product of the number of excited electrons and the number of vacancies. Since the number of excited electrons and vacancies is equal, the rate of recombination is given by:

$$\frac{dN}{dt} = -\zeta N^2, \tag{2.12}$$

and the number of excited centers at time t, which is obtained by integrating Eq. 2.12, is:

$$N = \frac{N_0}{1 + N_0\zeta t}. \tag{2.13}$$

Eqs. 2.12 and 2.13 being combined, the phosphorescent decay characteristic is found to be:

$$L = E\frac{1}{(1 + \gamma t)^2}. \tag{2.14}$$

The equilibrium between the transitions of excitation and emission being determined, the following relation between L and ρ can be obtained:

$$L = F\rho\{1 + \rho\delta - \sqrt{\rho^2\delta^2 + 2\rho\delta}\}. \tag{2.15}$$

Both these types of decay and current characteristics are found among the phosphors. Zinc sulphide unmistakably belongs to the second type. Sodium chloride activated with thallium is a member of the first group.

2.7. Phosphors for Television. A list of some of the principal cathodoluminescent phosphors, together with their properties, is given in Table 2.2. Of these, a number are suitable for viewing-tube screens, meeting to an adequate degree the conditions outlined in section 2.2.

The color requirement is, in general, easily satisfied. Psychologically,

TABLE 2.2

PRINCIPAL PHOSPHORS USED FOR CATHODOLUMINESCENCE

No.	Phosphor	Chemical Composition	Color	Spectral Maximum (Å)	Approximate Candle-power per Watt*
1	Zinc oxide	ZnO	Violet	Ultra-violet	<0.1
2	Zinc sulphide	$ZnS{:}Ag$	Blue-violet	4700–4500	5_T
3	Calcium tungstate	$CaWO_4$	Blue	4300	<1
4	Zinc silicate	$ZnO + SiO_2$	Blue	4200	<1
5	Zinc sulphide	ZnS	Light blue	4700	$1\text{–}5_T$
6	Zinc aluminate	$(ZnO + Al_2O_3){:}Mn$	Green-blue	5130	~1
7	Zinc silicate (willemite)	$(ZnO + SiO_2){:}Mn$	Blue-green	5230	3_T
8	Zinc sulphide	$ZnS{:}Cu$	Green	4700–5250	$>4_T$
9	Zinc germanate	$(ZnO + GeO_2){:}Mn$	Yellow-green	5370	1.5
10	Beta zinc silicate	$(ZnO + SiO_2){:}Mn$	Green-yellow	5600–5700	3_T
11	Zinc beryllium silicate	$(ZnO + BeO + SiO_2){:}Mn$	Green to orange	5230–6500	$1\text{–}2_T$
12	Zinc cadmium sulphide	$(ZnS + CdS){:}Ag$	Blue to red	4700–>7000	5_T
13	Calcium silicate	$(CaO + SiO_2){:}Mn$	Green to orange	5500–6500	<1
14	Cadmium silicate	$(CdO + SiO_2){:}Mn$	Orange yellow	5850	~1
15	Magnesium silicate	$(MgO + SiO_2){:}Mn$	Orange-red	6400–6700	<1
16	Zinc aluminate	$(ZnO + Al_2O_3){:}Cr$	Red	>7000	<1
17	Zinc beryllium zirconium silicate	$[ZnO + BeO + (Ti - Zr - Th - O_2) + SiO_2]{:}Mn$	White	4200 + 5500–6000	$\sim 1_T$
18	Magnesium tungstate	$MgO + WO_3$	Very light blue	4800	<1
19	Zinc borate	$(ZnO + B_2O_3){:}Mn$	Yellow-orange	5400–6000	$\sim 1_T$
20	Cadmium borate	$(CdO + B_2O_3){:}Mn$	Green-orange	5300–6300	<1
21	Cadmium tungstate	$CdO + WO_3$	Light blue	4900	$<1_T$

* T = used in television.

the color which produces the least fatigue is a yellow-green. However, perhaps because of long familiarity with photographs and other black-and-white reproductions, most observers are found to prefer a white screen to all others. As yet no efficient single phosphor is known which produces a white luminescence. However, by mixing two or more phosphors a white material can be produced. Thus, a green-blue zinc sulphide mixed in proper proportions with a yellow zinc cadmium sulphide, or zinc beryllium silicate, forms a screen material which gives white luminescence. It is interesting to note that the spectral intensity distribution shown in Fig. 2.8 of this "white" is very different from the "white" of daylight, which is nearly uniform over the entire visible band. This type of distribution produces the sensation of white because of the tricolor nature of human vision.

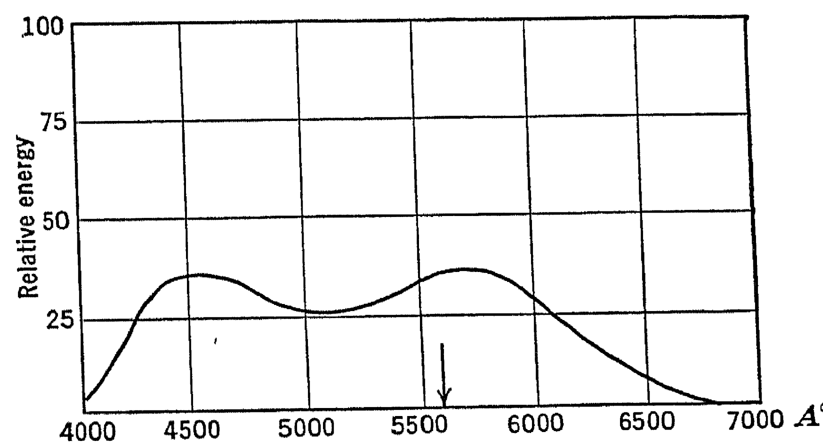


Fig. 2.8.—The Spectral Characteristics of a White Fluorescent Screen Consisting of Silver-Activated Zinc Sulphide and Zinc Cadmium Sulphide.

In general, the practical phosphors can be divided into two classes: namely, the sulphide phosphors and the oxide phosphors. Of the first group, zinc sulphide and zinc cadmium sulphide are most frequently used. Zinc orthosilicate and zinc beryllium silicate, of the second group, also find wide application.

2.8. Sulphide Phosphors. One of the most efficient cathodoluminescent materials known is zinc sulphide. This material when activated with silver has a luminous output of 3 to 5 candlepower per watt. As often happens, this material is superior in one respect but not so good in certain others. Perhaps the most important defect is its relative instability. However, recent improvements in the technique of applying the material to television tubes and in tube exhaust have resulted in screens which have a life nearly equal to that of the most durable material, and long enough so that the life of the tube is not determined by the life of the screen.

There would be no advantage in taking up the details of a factory

process of preparing luminescent zinc sulphide; however, it is not out of place to describe the laboratory method of synthesizing the phosphor.

Zinc sulphate is first purified by ordinary methods until the spectroscope fails to reveal any impurities. Then, in order to make certain that no copper remains, an aqueous solution of the salt is electrolyzed using platinum electrodes. Pure white zinc sulphide is precipitated from the purified solution by bubbling clean hydrogen sulphide through it. The precipitate is washed with the purest distilled water and dried. Ten grams of this zinc sulphide are placed in an acid-cleaned quartz crucible. After the addition of 0.2 gram of pure sodium chloride, the mass is ground thoroughly with a quartz rod. To this is added a silver nitrate solution which contains 0.001 gram of silver. The material is then evaporated to dryness and thoroughly ground. The phosphor is crystallized by heating

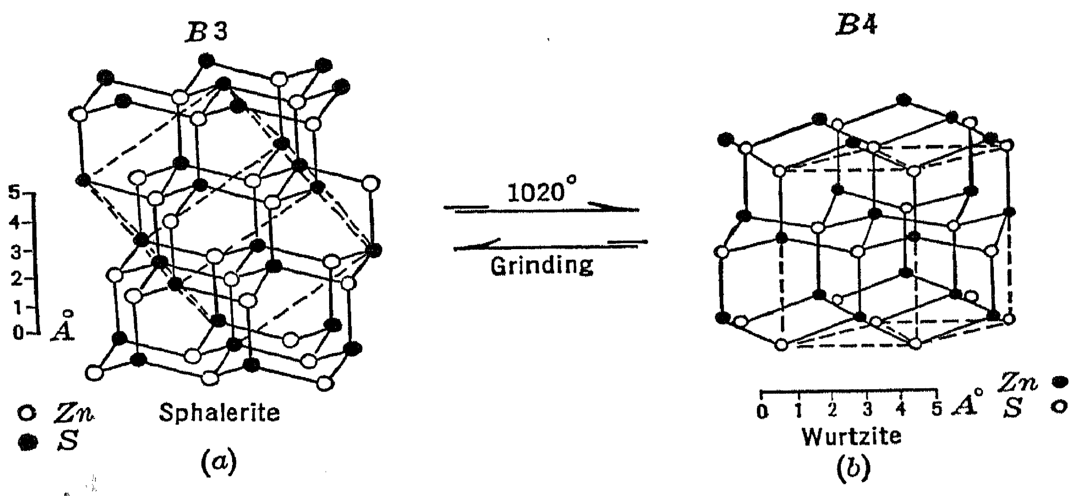


FIG. 2.9.—The Crystal Structure of Sphalerite and Wurtzite.

to 900°C, the temperature not being at all critical. The result will be ZnS:Ag, one of the best phosphors known to date.

The crystal structure of this material, known as sphalerite, is shown in Fig. 2.9*a*. There is also an enantiotropic form of this material, wurtzite, which crystallizes at temperatures above 1020°C, and which can be converted to sphalerite by grinding. The structure of this form is shown in Fig. 2.9*b*.

In spite of its high luminous efficiency, the phosphorescent decay period of ZnS:Ag is quite short, amply so for television purposes. Its persistence characteristics are given in Fig. 2.10. From this curve it will be seen that the light output has dropped to about 0.2 per cent in one picture period.

The light output is not strictly proportional to the bombarding current and tends toward a gradual saturation, as illustrated in Fig. 2.11. The variation with voltage is less readily obtained, as will be explained later,

because of the difficulty of determining the true bombarding voltage. In general, the light output increases with nearly the second power of the bombarding voltage.

The color of the luminescence is blue, corresponding to a band extending from 4000Å to 5700Å. A spectral curve (1) of this material is shown

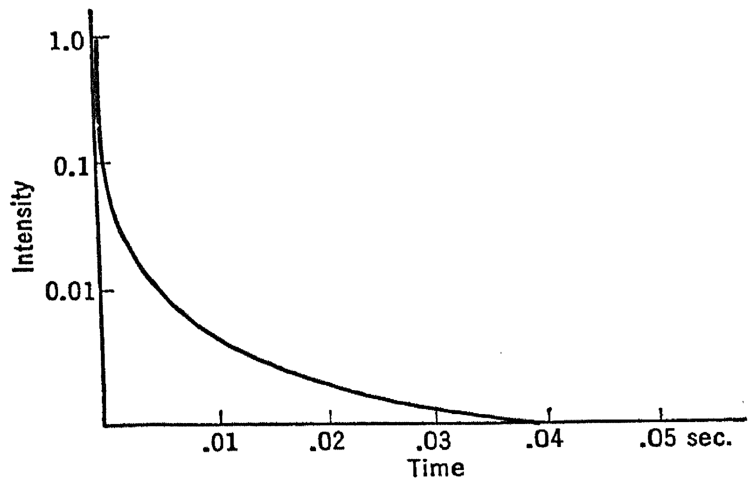


Fig. 2.10.—Persistence Characteristics of Zinc Sulphide.

in Fig. 2.12. However, this curve is merely representative, as the exact shape of the curve is very sensitive to composition, preparation, and after-treatment.

Zinc cadmium sulphide is another useful sulphide phosphor. The preparation is quite similar, except that, for every gram of purified zinc sul-

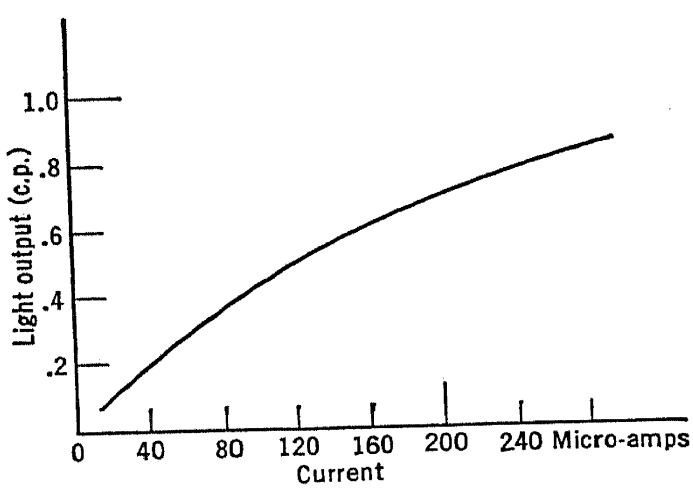


Fig. 2.11.—Luminescence as a Function of Electron Current for Zinc Sulphide.

phide, 1 gram of cadmium sulphide is added. Sodium chloride is used as a flux; 0.02 per cent silver in the form of the nitrate, as activator.

From the spectral output curve (2) of Fig. 2.12, the color of this material can be seen to be a yellow. Its luminous efficiency is higher than

that of the zinc sulphide. The persistence characteristics of the two sulphide phosphors are very similar.

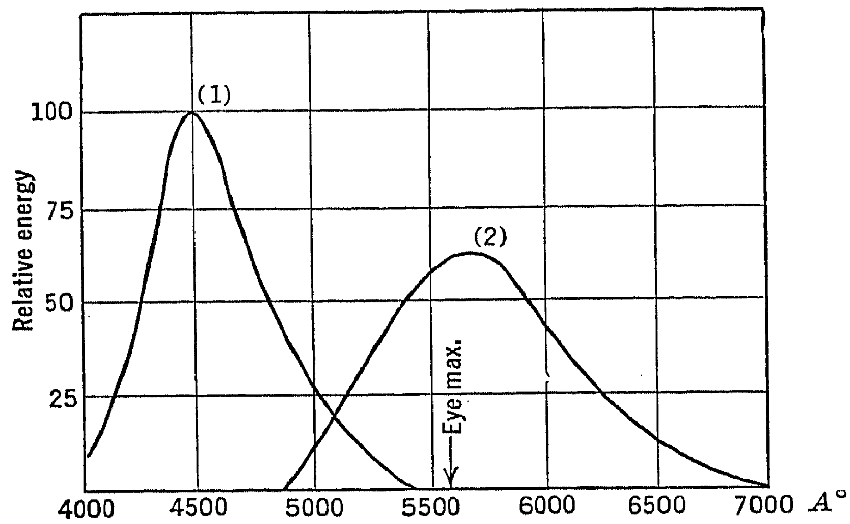


Fig. 2.12.—The Spectral Characteristics of Silver-Activated Zinc Sulphide and Zinc Cadmium Sulphide.

2.9. The Oxide Phosphors. The best-known and most-used representative of this class is artificial willemite, or zinc orthosilicate. This material, though not so efficient as the two sulphides just described, has a light output of 3 candlepower per watt. Its color can be made to range from green to pure yellow with the same high efficiency, and can be varied over the entire visual range at the expense of efficiency.

The main advantage of this phosphor over those previously discussed is its extreme ruggedness, which is no small advantage in television work, particularly in its experimental stages.

Small-scale laboratory preparation of this material can be carried out as follows.

Twenty grams of highly purified and finely divided zinc oxide and 7.4 grams of pure silicon oxide are ground together thoroughly. To this is added about one mole per cent of manganese in solution, or 0.07 gram of the metal. The mixture is stirred, dried, and then finely ground. Lastly, it is heated in a covered platinum crucible to about 1300°C for two hours. If the heated material is cooled slowly the resultant phosphor has a bright green luminescence; if it is quenched rapidly from a melt (above 1512°C) the luminescence will be yellow.

The crystal structure of this material is much more complicated than that of the sulphide. The α form, that is, the material which is formed by a gradual cooling, has a structure which is isomorphous with phenacite, with 6 molecules to the unit cell. The second, or β form, has not been completely worked out, although preliminary X-ray measurements have been made. The material may exist in a third, or γ, state when pre-

pared with an excess of silica, but this is of little interest for our present purpose as it exhibits only a weak red luminescence.

The persistence characteristics for the two forms of $(ZnO + SiO_2)$:Mn are shown in Fig. 2.14. In spite of the difference in the spectral output,

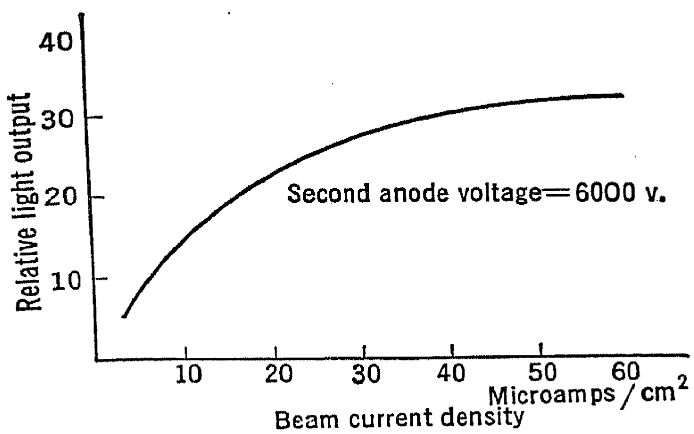


FIG. 2.13.—Luminescence as Function of Electron Current for Willemite.

the decay curves are very similar. The luminous output of both these materials has dropped to about 4 per cent of its initial value in one picture period.

The spectral characteristics of the two materials can be seen from Fig. 2.15. These phosphors give the characteristic broad band of radiation.

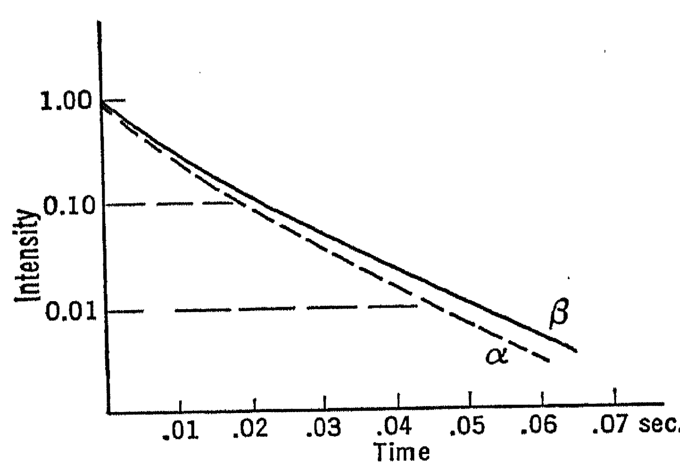


FIG. 2.14.—Persistence Characteristics of α and β Zinc Orthosilicate.

Like the sulphides, these materials exhibit a gradual saturation with current, and a light output which rises somewhat faster than with the first power of the voltage.

Another specific phosphor that might be mentioned because of its

practical value is calcium tungstate. This is a member of the rather restricted class of phosphors which either emit without the aid of a foreign activator, or cannot be purified sufficiently to remove the requisite impurity. Calcium tungstate has a rather low efficiency, i.e., about 0.3 lumen per watt, and luminesces with a pale blue color. However, its decay is very rapid, the light output dropping to 10 per cent in something like 10^{-7} second. This makes it valuable for oscillograph tubes designed for uses where long decay cannot be tolerated. Its emission is highly actinic, which is desirable for photographic purposes. Zinc sulphide activated with a minute quantity of nickel similarly has a phosphorescence of extremely short duration.

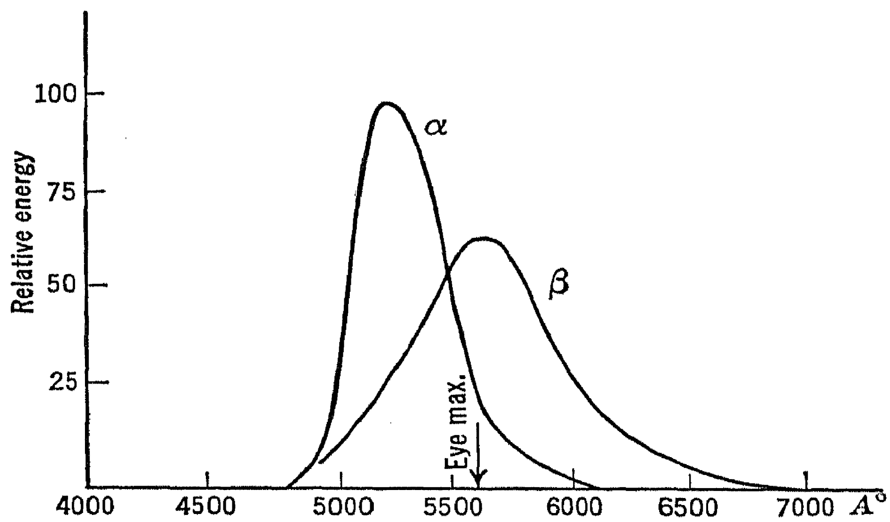


FIG. 2.15.—Spectrum of Luminescence of α and β Zinc Orthosilicate.

2.10. Electrical Properties of Phosphors. The specific resistance of all materials used as phosphors is extremely high. Therefore, when a phosphor is bombarded by an electron beam of the current density used in television practice, the principal way the electrons escape from the surface is by secondary emission. If the secondary-emission ratio of the surface is less than unity for a given bombarding voltage, the area will become more negative until it either reaches cathode potential, thus preventing electrons from reaching the screen, or else reaches a potential such that the secondary-emission ratio becomes unity.

Because of its practical importance it is desirable to discuss the discharge process in some detail. As has been shown in the preceding chapter, the secondary-emission ratio of most substances, conductors and insulators alike, rises rapidly to a maximum in the neighborhood of 300 to 800 volts, and then falls slowly as the voltage increases. This is true of all the phosphors that have been studied, and also of the glasses which are used as a backing to support the screen material. The maxima and shape of the secondary-emission curves are characteristic of each sub-

stance. However, the forms of the curves are sufficiently similar to permit useful generalizations. Fig. 2.16 shows a typical secondary-emission curve for an insulator.

Another factor determining the rate at which the electrons leave the screen is the potential difference between the bombarded surface and the electrode which collects the current from it. In order to obtain a maxi-

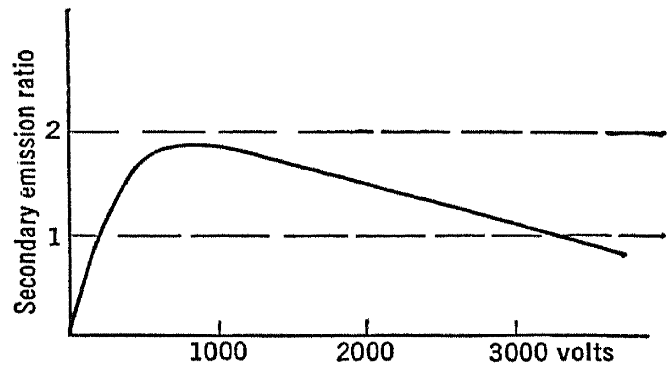


Fig. 2.16.—Typical Secondary Emission Curve for an Insulator.

mum current, it is necessary that the collector be at a positive potential with respect to the screen; however, current will continue to flow to the electrode even when it is slightly negative, owing to the initial velocities of the leaving electrons. The form of the curve representing the relationship between collector potential and current ratio is illustrated in Fig. 2.17. The shape of this curve is determined by such factors as tube con-

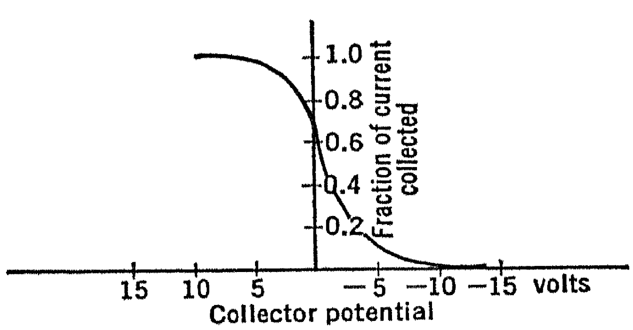


Fig. 2.17.—Secondary Emission Current as Function of Collector Voltage.

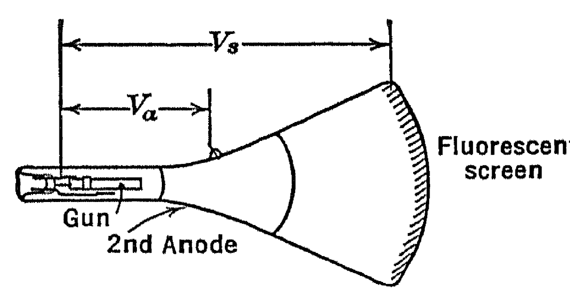


Fig. 2.18.—Schematic Diagram of a Typical Kinescope.

figuration, bombarding current, and the nature of the emitter. However, the range of potentials represented by the rising portion of the saturation curve is so small compared with the bombarding voltage that the use of a generalized form does not appreciably affect the validity of the conclusions drawn.

The type of tube in which the screen material is used is shown in Fig. 2.18. The figure illustrates schematically a typical Kinescope which con-

sists of an electron gun acting as the cathode-ray source, a second anode which collects the electrons after they leave the screen, and a screen of luminescent material. The voltage V_a is applied between the gun and second anode. However, since the additional condition that as much current must leave the screen as arrives in the beam must be fulfilled, the velocity with which the electrons strike the screen will differ, in general, from V_a volts. On the diagram the true bombarding voltage is represented as V_s, which normally is less than V_a but in certain cases may rise above this value.

The surface of the screen is thus bombarded by electrons with velocity V_s, and is subject to a collector potential $V_a - V_s$, the only parameter

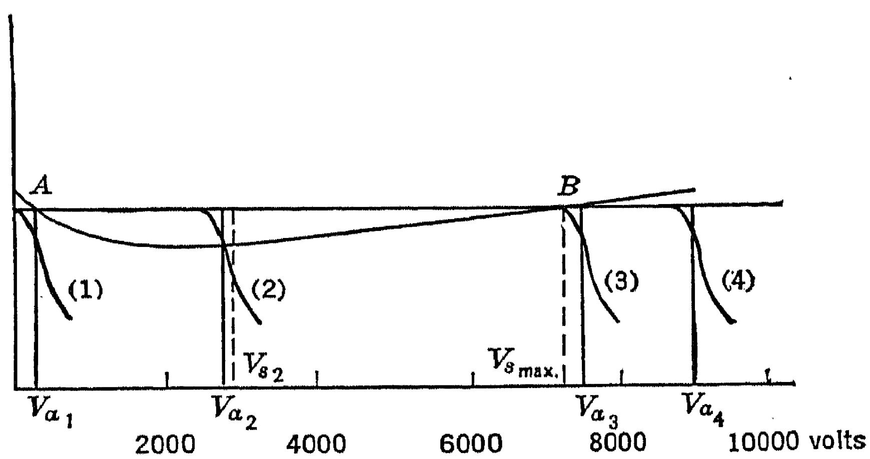


Fig. 2.19.—Equilibrium Conditions of a Bombarded Insulator.

being V_a. Let it be assumed that the function $S(V_s)$, giving the secondary-emission ratio in terms of the true bombarding voltage, and $C(V_a - V_s)$, describing the variation of the collector current with the collector voltage, are known. The latter function, $C(V_a - V_s)$, is the ratio of the current actually collected to the saturated secondary emission at the voltage V_s in question. The current collected will be:

$$i = i_b\, S(V_s)\; C(V_a - V_s). \tag{2.16}$$

Since, for equilibrium, the beam current i_b must equal the collector current, the relation:

$$\frac{1}{S(V_s)} = C(V_a - V_s) \tag{2.17}$$

is necessarily fulfilled. If a plot is made of the inverse secondary-emission function, and of the saturation curve with its origin at the second anode voltage V_a, the intersection of the two curves will be at the true bombarding voltage V_s. Fig. 2.19 is an example of such a plot showing the determination of V_s for several values of the second anode potential V_a.

It is instructive to consider the behavior of the screen as shown by

this diagram. For second anode voltages from 0 to A, the bombarding voltage is zero; in other words, the beam is turned back and does not strike the screen. Between A and B, where the secondary-emission ratio is greater than unity, the screen voltage is approximately that of the second anode and may actually be more positive by one or two volts. As the second anode voltage approaches B, the screen voltage drops below the second anode voltage. Beyond B, the screen voltage stays constant at the value B regardless of the second anode voltage. Thus, a curve of the true bombarding voltage as a function of second anode voltage has the form shown in Fig. 2.20. The break-point occurs at a different second

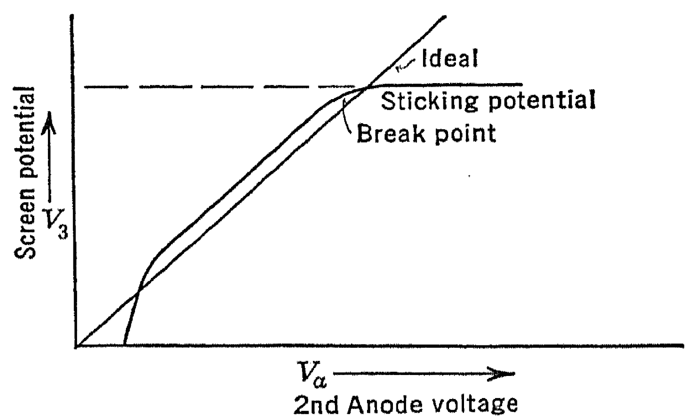


FIG. 2.20.—Potential of a Fluorescent Screen Relative to the Electron Source as Function of Second Anode Voltage.

anode voltage, depending upon the material used as screen and the treatment to which it has been subjected.

Measurements have been made of the break-point or sticking potential for glass, willemite, and zinc sulphide.* For Pyrex, such as is used in the manufacture of Kinescopes, the break-point is found to lie between 2000 and 3000 volts. For willemite, it varies between 5000 and 8000 volts, depending upon the exact preparation. Only a few measurements are at present available on zinc sulphide, and these indicate the same range of values as for the orthosilicate.

When these materials are applied in an actual tube, conditions are much more complicated and a much wider range of values for the break-point is found. In this case such factors as screen thickness, getter, residual gas, and age are important.

The variation of the bombarding voltage with the age of the tube is indicated in Fig. 2.21. This change may be due to the getter material, which greatly increases the secondary-emission ratio of the phosphor, being driven from the screen, or to small amounts of gas, etc., being freed from the walls or the gun.

* W. B. Nottingham, reference 9; H. Nelson, reference 12.

Even more interesting is the apparent variation of break-point with screen thickness. Referring to Fig. 2.22, it will be seen that for very thin screens the break-point is low, corresponding to that of the glass, and that for very thick screens it is similar to that of pure willemite. For in-

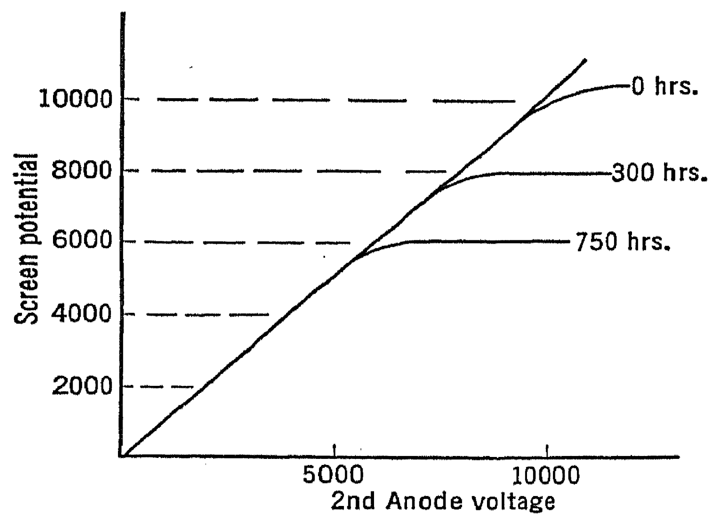


FIG. 2.21.—Variation of Screen Potential Characteristics with Tube Age.

termediate screen thickness the break-point is no longer sharp, and the screen and second anode potentials do not begin to differ materially until very high voltages are reached. In the very thin screen every particle of the phosphor is surrounded by a region of exposed glass which prevents

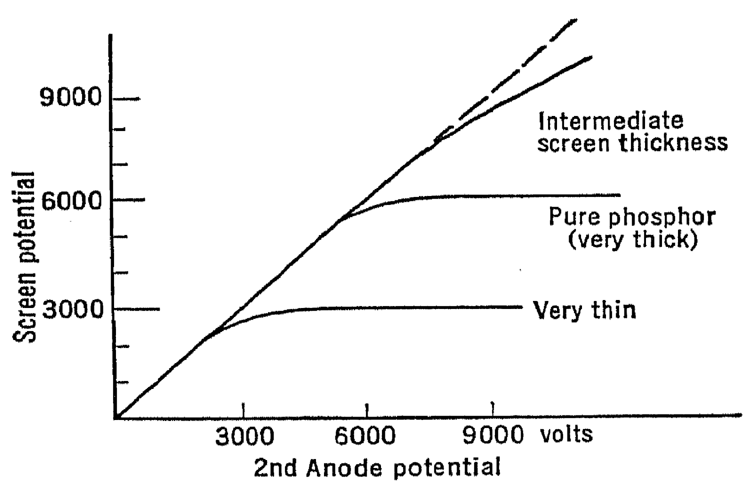


FIG. 2.22.—Variation of Screen Potential Characteristics with Screen Thickness.

its potential from rising much above the break-point of glass by its grid action on the low-velocity secondary electrons from the phosphor. The break-point of the thick screen is just that of the pure phosphor itself. The conditions for the intermediate screen are more complicated, and only a conjectural explanation of its behavior can be given. There is not enough exposed glass to produce the grid action as in the thin screen,

yet bombarding electrons strike both glass and phosphor. If these substances were to reach their characteristic break-points, it would mean that a potential of several thousand volts would exist in distances comparable to the screen thickness, which is only a few microns. Such a gradient may be sufficient to produce a breakdown which serves to discharge both phosphor and glass, and prevents establishing a true break-point.

Orthosilicate and sulphide screens for use in viewing tubes can be prepared in such a way that the screen potential continues to rise with second anode voltage at potentials far above the break-point of the pure phosphor. Suitably processed screens can be used at voltages even in excess of 50 kilovolts.

The scanning operation adds a further complication to the picture of the electrical behavior of the screen in a tube used for television purposes. Measurements under these conditions show that, above the break-point, the screen potential tends to be closer to second anode potential at points not under the beam than at the spot directly under the beam. Where the second anode is 3000 volts above the break-point, a potential variation of as much as 1000 volts may exist over the screen.

If television development were going to stop with the direct-vision Kinescope, which operates in general below the break-point, the above discussion would be quite valueless. However, there is every indication that the projection tubes operating at extremely high voltages to obtain maximum brilliancy from the phosphor will be of the utmost importance in the not too distant future. For these tubes, considerations of secondary emission and other discharge mechanisms will become quite essential.

2.11. Conclusion. No mention has so far been made of methods of applying a phosphor to screens used in television viewing tubes. These, together with the related problems of optimum particle size, screen thickness, factors influencing contrast, etc., are reserved for a later chapter dealing specifically with the Kinescope, since they pertain to a particular use of the material.

As to the future development of better luminescent materials, there is every reason to be optimistic in view of the progress that has been made in the past few years. That great improvement in efficiency may be possible is evident from a consideration of present efficiencies and the efficiency that can be obtained through ultra-violet excitation. The conversion of 1 watt of energy into radiation producing the maximum visual response will give rise to about 620 lumens of light and, if converted into radiation with the spectral characteristics of willemite, approximately 500 lumens. The cathodoluminescence of willemite yields only 35 or 40 lumens per watt, or less than 10 per cent efficiency. On the other hand,

it has been reported that ultra-violet excitation of this same material in a mercury-vapor fluorescent lamp will give as much as 245 lumens per watt of ultra-violet energy or, in other words, nearly 50 per cent efficiency. There is no known theoretical reason for believing that efficiencies equal to, or greater than, those obtained with ultra-violet excitation cannot be obtained for cathodoluminescent phosphors. It is to be expected not only that the efficiency of the material will be markedly increased, but also that its performance at high currents and voltages, the color of its fluorescence, and its durability will be improved.

REFERENCES

1. P. Lenard, F. Schmidt, and R. Tomaschek, "Phosphorescence and Fluorescence," *Handbuch der Experimentalphysik*, Vol. 23, Parts 1 and 2, Leipzig, 1928.
2. H. Rupp, "Luminescent Materials and Their Application," Gebrüder Borntraeger, Berlin, 1937.
3. J. T. Randall, "Luminescence and Its Applications," Royal Society of Arts, London, 1937.
4. L. Levy and D. W. West, "Fluorescent Screens for Cathode-Ray Tubes for Television and Other Purposes," *J.I.E.E.*, Vol. 79, pp. 11–19, July, 1936.
5. T. B. Perkins and H. W. Kaufman, "Luminescent Materials for Cathode Ray Tubes," *Proc. I. R. E.*, Vol. 23, pp. 1324–1333, November, 1935.
6. F. Seitz and R. P. Johnson, "Theory of Solids," *J. Applied Phys.*, Vol. 8, pp. 84–97, February, 1937; pp. 186–199, March, 1937; pp. 246–260, April, 1937.
7. F. Seitz, "Alkali Halide-Thallium Phosphors," *J. Chem. Phys.*, pp. 150–162, March, 1938.
8. F. Seitz, "Interpretation of the Properties of Zinc Sulphide Phosphors," *J. Chem. Phys.*, pp. 454–461, August, 1938.
9. W. B. Nottingham, "Electrical and Luminescent Properties of Willemite under Electron Bombardment," *J. Applied Phys.*, Vol. 8, pp. 762–778, November, 1937.
10. S. T. Martin and L. B. Headrick, "Light Output and Secondary Emission Characteristics of Luminescent Material," *J. Applied Phys.*, Vol. 10, pp. 116–127, February, 1939.
11. E. G. Ramberg and G. A. Morton, "Variation of Light Output with Current Density and Classification of Willemite," *Phys. Rev.*, Vol. 55, February, 1939.
12. H. Nelson, "Method of Measuring Luminescent Screen Potentials," *J. Applied Phys.*, Vol. 9, pp. 592–599, September, 1938.
13. F. Seitz and H. W. Leverenz, "Luminescent Materials," *J. Applied Phys.*, Vol. 10, pp. 479–493, July, 1939.

CHAPTER 3

ELECTRON OPTICS

The term "electron optics" will be used in the present chapter to describe that class of problems which deals with the determination of electron trajectories. The expression originated as a consequence of the close analogy between optical arrangements and the corresponding electronic systems. It was found that this analogy not only had fundamental mathematical significance, but, in many cases, could be extended to practical devices. For example, it is possible to construct electron lenses which are capable of imaging an electron source. In many instances not only is the behavior of the two types of systems the same, but also many of the mathematical methods of optics can be applied to the corresponding electron problem. There are, it is true, many systems which in no way resemble those of conventional optics. However, there is a continuous transition between these and such as have a close optical analogue. Therefore, any attempt to subdivide the field on this basis results only in confusion.

The importance of electron optics is becoming increasingly apparent with the advance of electronics. For example, in the early vacuum tubes used in radio work little attention was paid to the exact paths of the electrons between the cathode and the plate. Recently, very real improvement in efficiency and performance has been achieved by the application of electron optics to tube design. In the design of the newer devices, such as the secondary-emission multiplier, the electron gun used in television tubes, and the image tubes, electron optics is essential.

The design problem usually encountered is one in which the two termini of the electron paths are specified and it is required to determine an electrode and magnetic coil configuration that will satisfy this demand. Unfortunately, a direct solution is still a good deal beyond our present mathematical means. It is not possible, except in very special cases, to determine from a given electron path the shape and potentials of the electrodes required to produce this path. In order to solve the above problem it is necessary to assume an electrode configuration and then determine the resulting electron path. If this is not the required path the electrodes are changed and the trajectories recalculated. Usually this process does not have to be repeated very many times before the

correct solution is reached, as the previously determined paths indicate the nature of the changes that must be made.

Restating the problem in the only form in which a solution is possible, it becomes: *Given an electrode configuration and the potentials applied, determine the electron paths in the resulting potential field.*

Even this problem has no general solution, and often can be solved only by resorting to elaborate mathematical approximations, or to the use of mechanical and graphical methods. The solution can be divided into two distinct parts: namely, the determination of the potential field produced by the electrodes, and the calculation of the electron trajectories in this field. Essentially the same procedure is used when the electrons are guided by magnetic fields.

3.1. The Laplace Equation. To determine the potential field produced by a given set of electrodes, it is necessary to solve the Laplace differential equation:

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0, \tag{3.1}$$

with boundary conditions corresponding to the shapes and potentials of the electrodes. The solution of this equation gives the potential as a function of the coordinates, that is:

$$\phi(x,y,z).$$

The electrostatic field can be found from this potential by differentiation with respect to the coordinates. Thus:

$$E_x = -\frac{\partial}{\partial x}\,\phi(x,y,z);$$

$$E_y = -\frac{\partial}{\partial y}\,\phi(x,y,z);$$

$$E_z = -\frac{\partial}{\partial z}\,\phi(x,y,z).$$

From the original Laplace equation, which is satisfied by the potential function, it will be seen that the field must satisfy the differential equation:

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0.$$

It should be noticed that this equation is like an equation of continuity

and may be interpreted to mean that in any volume of free space within an electrode system as many electrostatic lines of force must leave as enter. Similar equations express corresponding laws obeyed by the flow of an incompressible fluid and by electric current in a conducting medium.

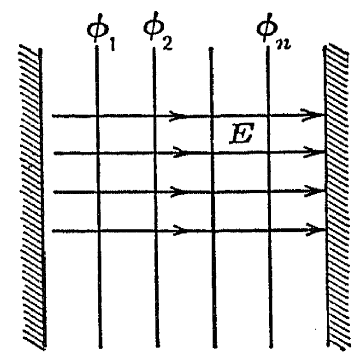


FIG. 3.1.—Potential and Field between Plates.

There is no general solution for the Laplace equation, nor can any general method of attack be given. In certain special cases only can an analytic solution be obtained. Usually it is necessary to resort to series expansion or numerical integration in order to calculate a potential distribution. Both procedures are laborious in the extreme.

The simplest potential distribution is that between two infinite parallel plates, shown in Fig. 3.1. Here the function that satisfies the differential equation is:

$$\phi = \phi_0 + Ax.$$

The field will be found to be:

$$E_x = -\frac{\partial \phi}{\partial x} = -A,$$

$$E_y = E_z = -\frac{\partial \phi}{\partial y} = 0.$$

Other simple cases are:

Concentric spheres: $$\phi = -\frac{A}{r} + \phi_0;$$

$$E_r = -\frac{A}{r^2}.$$

Concentric cylinders: $$\phi = A \ln r + \phi_0;$$

$$E_r = -\frac{A}{r}.$$

Other cases, such as two separated spheres, a sphere and plane, a sphere between two planes, and the corresponding cylindrical systems, can also be solved.

Problems involving cylindrical symmetry, such as illustrated in Fig. 3.2, are of considerable importance, since, as will be shown in the next chapter, this symmetry is found in all electron lenses. When this sym-

metry is present the Laplace equation is preferably expressed in cylindrical coordinates, becoming:

$$\frac{\partial^2\phi}{\partial z^2} + \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right) = 0. \tag{3.2}$$

For most boundary conditions the solution of this differential equation is difficult, and no analytic solution is possible. A general method of

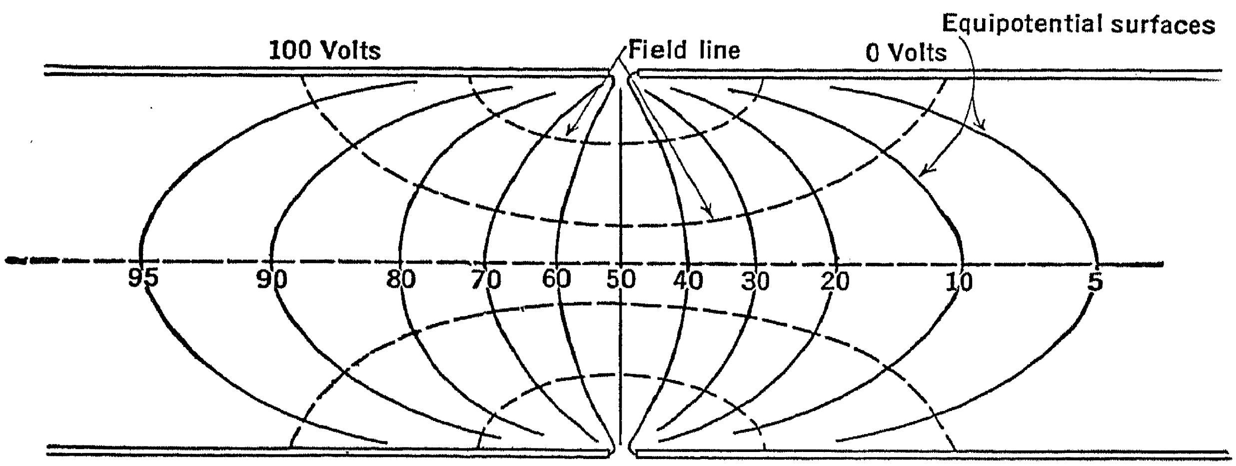


Fig. 3.2.—Potential Distribution between Two Cylinders.

attack is to consider the potential as a linear combination of functions in which the variables have been separated, thus:

$$\phi(r,z) = \phi_1 + \phi_2 + \cdot \cdot \cdot + \phi_k + \cdot \cdot \cdot,$$

where:

$$\phi_k(r,z) = F_k(z)G_k(r). \tag{3.3}$$

By substituting Eq. 3.3 in Eq. 3.2, the Laplace equation reduces to two ordinary differential equations:

$$\frac{1}{F}\frac{d^2F}{dz^2} = -k^2,$$

$$\frac{1}{rG}\frac{d}{dr}\left(r\frac{dG}{dr}\right) = k^2,$$

where k^2 is the separation parameter. The general solution of these equations can be written as:

$$F_k(z) = ae^{ikz} + be^{-ikz}, \tag{3.4a}$$

$$G_k(r) = cJ_0(ikr) + dN_0(ikr). \tag{3.4b}$$

The solution of the Laplace equation then has the form:

$$\phi(r,z) = \sum_k A_k F_k(z)G_k(r). \tag{3.5}$$

Since k does not necessarily have discrete values, Eq. 3.5 may take the form of an integral:

$$\phi(r,z) = \int A(k)F(z,k)G(r,k)dk, \tag{3.5a}$$

the integration being over the entire complex domain. The coefficient $A(k)$ is determined from the boundary conditions by the usual methods of evaluating Fourier coefficients.

Another class of problem of considerable importance is that in which the potential expressed in Cartesian coordinates is a function of two of these coordinates only. This type of potential field is encountered wherever the electrode surfaces can be considered as generated by moving lines which remain parallel.

The Laplace equation in this case becomes:

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} = 0. \tag{3.6}$$

The solution of this equation can be approached in a variety of ways. One very useful method, for example, is that of conformal mapping. Although this equation can be solved more frequently than that for the three-dimensional case, often no analytic solution is possible. In this two-dimensional case, practical electrode configurations are usually quite complicated, so that the mathematical complexities even of an approximate solution are prohibitive.

3.2. Electrolytic Potential Mapping. Because of the difficulties encountered in a mathematical solution of the Laplace equation, it is often expedient to resort to an electrolytic method of obtaining an equipotential map.

In essence, the method consists of immersing a large-scale model of the electrode system being studied in a slightly conducting liquid and measuring the potential distribution throughout the liquid with a probe and bridge, potentials proportional to those to be used with the system being applied to the model. Fig. 3.3 shows diagrammatically an electrolytic plotting tank.

The tank used for this purpose is constructed of an insulating material so that the equipotential surfaces about the immersed electrodes meet the tank walls perpendicularly. This is necessary in order to reduce the influence of the walls of the tank upon the field to be plotted. The size of the tank is determined by the size of the electrode models which are to be studied, and this in turn is determined primarily by the accuracy desired. In order to reduce the boundary effects, the tank must be a good deal larger than the model.

The electrolyte used in the tank is water to which a very small amount of soluble salt has been added. In most localities, ordinary city water contains a sufficient amount of these salts to make it amply conducting for the purpose.

Exact scale models of the electrodes, made of sheet metal and supported on insulating rods, are used in the tank. The supports should be reduced to a minimum so that they interfere with the potential distribution as little as possible. Almost every practical electrode configuration encountered in electron optics has mirror symmetry. The equipotential surfaces in space around the electrodes obviously must cross the plane of symmetry at right angles. Because of this, as will become clear as the discussion proceeds, the models may be constructed so that they represent, to scale, half of each electrode bounded by the plane of symmetry. The model is placed in the tank in such a way that the free surface of the electrolyte coincides with the plane of symmetry. Although this is not a fundamental limitation, nearly all practical plotting tanks are limited to use with electrode systems having this symmetry.

Upon the application of the proper potentials to the model electrodes, a current will flow through the electrolyte. Since it can be assumed that the electrolyte is an ohmic conductor, the field strength at any point will be proportional to the current density. As was mentioned above, the electric current behaves like an incompressible fluid so that the current density and hence the field strength obey the equation of continuity (i.e., their divergence is zero in the absence of sources or sinks). This is merely another way of saying that the potential throughout the electrolyte obeys the Laplace equation. Thus the potential at any point in the liquid is proportional to the potential of a corresponding point in the actual electrode system.

The free surface of the electrolyte is a perfect insulating plane since no current can flow in the medium above the liquid. The equipotentials must intersect such a plane at right angles because there is zero vertical current flow, and hence the field vector normal to the surface is zero. For this reason it is possible to make use of models divided at their plane of symmetry.

The potential distribution over the plane of symmetry is measured by means of an exploring probe. This probe consists of a fine wire mounted so as to just break the surface of the liquid and is constrained to move in a horizontal plane. The potential of the probe is adjusted until zero current flows, and the potential is noted. This potential is that of the point where the probe touches the surface. For convenience, the probe is carried at the end of a pantograph linkage, so that the motion of the probe is reproduced by a stylus attached to the other end of the

linkage. This stylus, or mapping pencil, moves over a plotting table. The arrangement will be clarified by reference to Fig. 3.3.

A photograph of a typical plotting tank is reproduced in Fig. 3.4. The tank itself is made of wood, coated on the inside with roofing cement to render it water-tight and shielded outside with sheet copper. It is 2½ feet wide, 8 feet long, and 2½ feet deep. Along one side is a table on

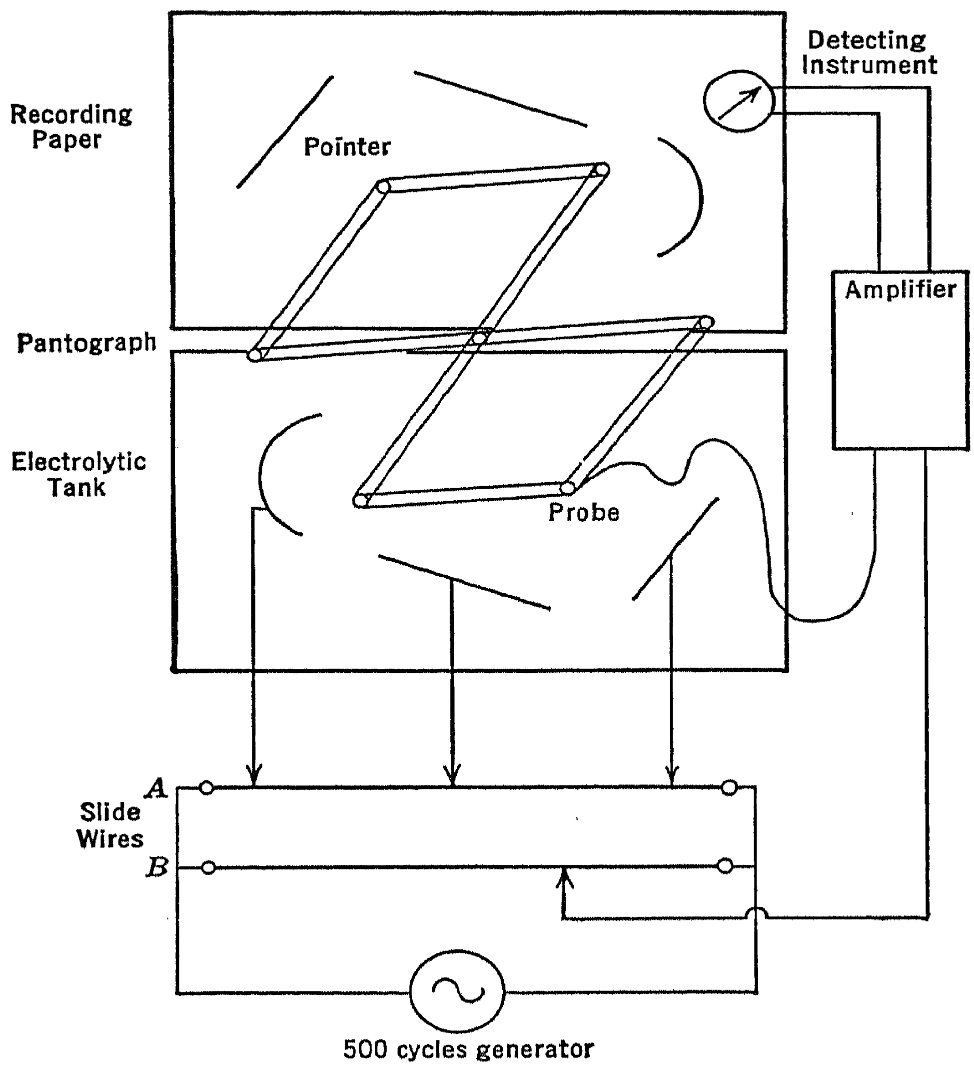


Fig. 3.3.—Diagram of Potential Plotting Tank.

which the mapping is done. Directly below the table are the potential dividers that supply the model electrodes and probe, and behind them the amplifier whose output is connected to the null-indicating meter. The probe is attached to the pantograph pivoted at the center of the edge of the tank nearest the mapping table. A shielded lead carries the current from the probe to the amplifier.

In the example illustrated in Fig. 3.4, the probe and electrodes are supplied from a 400-cycle oscillator, instead of with direct current. This not only facilitates the determination of the null point, but also avoids

the possibility of error due to polarization of the liquid at the probe or electrodes.

The usual procedure in mapping a potential distribution is to divide the voltage between the terminal electrodes into a convenient number of intervals, then to set the probe potential at each of these values in turn and map the path of the probe when it is moved in such a way that the current to it remains zero. The resulting map of the intersections of the equipotential surfaces corresponding to the voltage steps with the plane

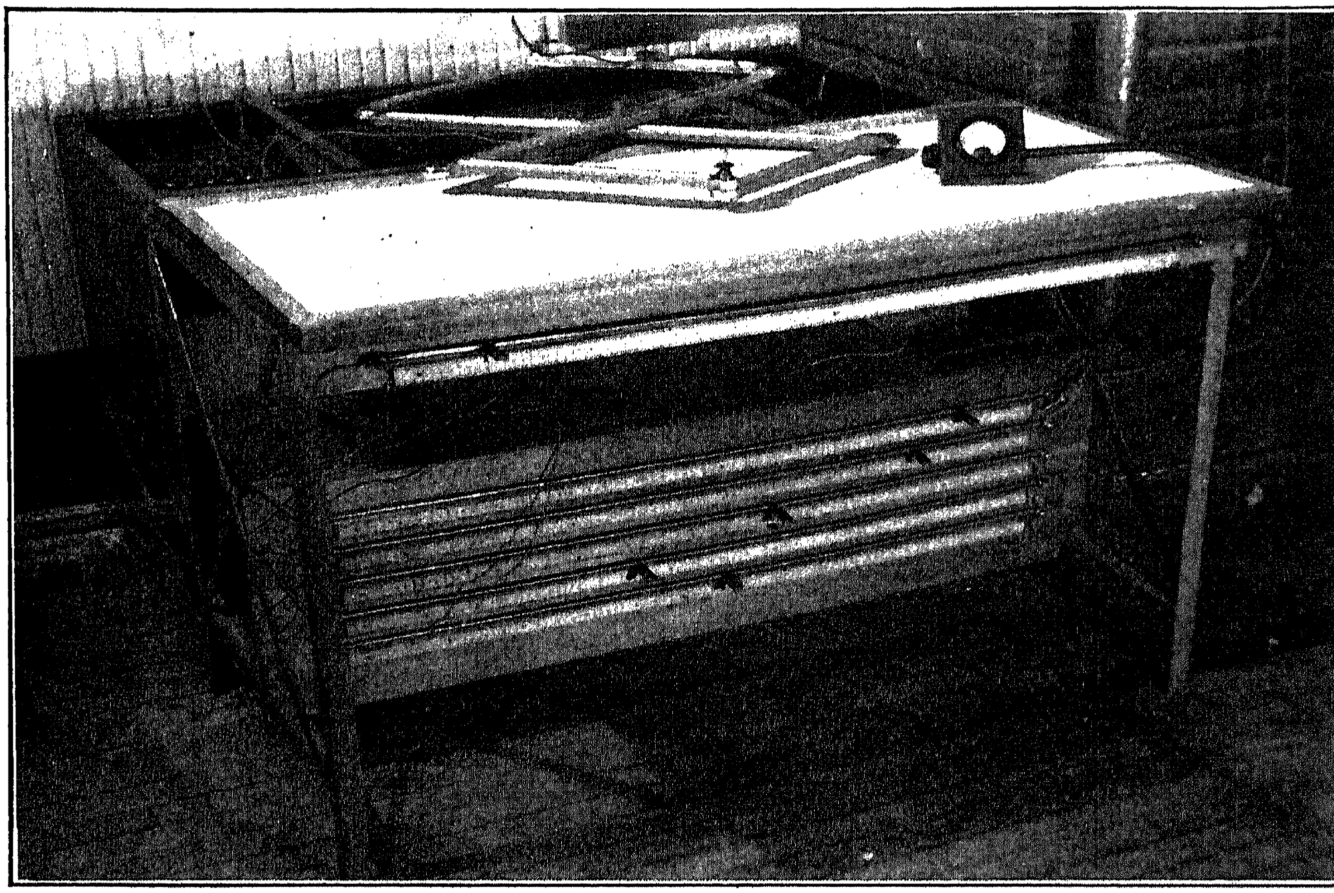


Fig. 3.4.—Potential Plotting Tank (RCA Manufacturing Company).

of symmetry of the electron-optical system is the most convenient representation of a potential distribution for the determination of electron trajectories.

Often, in the consideration of electron lenses, to be taken up in the next chapter, it is necessary to know the axial distribution of the potential of the system, together with its first and second derivatives along the axis. The distribution can, of course, be found by direct measurement with the plotting tank. The slope of the distribution curve, plotted as a function of the axial coordinate, will give the first derivative. The second derivative can be found from the slope of the first derivative curve, but this method of obtaining it is very inaccurate. A more accurate deter-

mination can be made by means of the radius of curvature ρ of the equipotentials at the axis, the first derivative ϕ', and the relation:

$$\phi''(z) = \frac{2\phi'}{\rho}.$$

3.3. The Motion of an Electron in a Potential Field. The potential distribution for a given electrode configuration having been obtained, the next step is the determination of the paths of electrons moving in this field.

In making this determination, it is convenient to consider the motion of an electron to be that of a charged particle of mass m obeying the laws of Newtonian mechanics, rather than to adopt the viewpoint of quantum physics and assume it to be a wave packet, as is necessary in the investigation of atomic phenomena. Furthermore, its mass will be taken as constant and equal to $9.0 \cdot 10^{-28}$ gram. Where electrons having extremely high velocities are to be considered, this assumption cannot be made, and it is necessary to correct for the increase in mass as dictated by relativity. Velocities where this correction is necessary are not encountered in the field covered by this book. The value 1.59×10^{-19} coulomb will be taken as the charge of an electron.

The force acting on an electron is the product of its charge and the field strength at the point which it occupies. Hence, the differential equations of motion are:

$$m\frac{d^2x}{dt^2} = -eE_x = e\frac{\partial\phi}{\partial x},$$

$$m\frac{d^2y}{dt^2} = -eE_y = e\frac{\partial\phi}{\partial y},$$

$$m\frac{d^2z}{dt^2} = -eE_z = e\frac{\partial\phi}{\partial z}.$$

In principle, in order to determine the electron path, all that is necessary is to introduce the values of the potential into the above equations, solve, and eliminate time as a parameter.

Actually, there is no general method of carrying out this process, and it is almost always necessary to apply mathematical approximations or graphical methods to obtain a solution.

A number of practical mathematical approximations and graphical methods have been developed for the purpose of facilitating the determination of electron paths when the potential field is known. These

methods, when carefully applied, are capable of yielding a high degree of accuracy.

The general three-dimensional problem is extremely difficult even by approximate methods. Fortunately, configurations requiring the solution of this general problem are rarely encountered, at least at the present time.

In the field covered by this book, two classes of problems are of particular importance. They are:

1. Those involving electrode configurations in which the potential variation is confined to a plane.
2. Problems involving cylindrical symmetry.

The remainder of this chapter will treat the first class of problems, which are those involved in the design of the electron multiplier, deflecting plates, etc.

3.4. Electron Paths in a Two-Dimensional System. Where the potential variation is confined to a plane, the Laplace equation, as has already been pointed out, involves two coordinates only:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0.$$

Similarly, the laws of motion become:

$$\left.\begin{aligned} \frac{d^2x}{dt^2} &= \frac{e}{m}\frac{\partial \phi}{\partial x}, \\ \frac{d^2y}{dt^2} &= \frac{e}{m}\frac{\partial \phi}{\partial y}. \end{aligned}\right\} \tag{3.7}$$

This particular form of these laws is not very convenient in the present consideration, in that time enters the equations explicitly.

Taking the principle of least action as a starting point simplifies the treatment, but it should be noted that all the relations deduced below can be derived directly from the force laws of Eq. 3.7.

The principle of least action states that any particle moving between two points in a potential field will follow a path such that the integral of the momentum over this path is an extreme, either maximum or minimum. Symbolically, this principle can be written as:

$$\delta \int_A^B mvds = 0. \tag{3.8}$$

An electron moving in a potential field has a kinetic energy just equal to the decrease in its potential energy during its motion. If the potential

is set equal to zero at a point where the electron is at rest, the following relation applies:

$$\tfrac{1}{2}mv^2 = e\phi,$$

or:

$$v = \sqrt{2\frac{e}{m}\phi}.$$

The momentum in the action integral being represented by $\sqrt{2em\phi}$, Eq. 3.8 becomes:

$$\delta \int_A^B \sqrt{\phi}\, ds = 0, \tag{3.9}$$

or:

$$\delta \int_A^B \sqrt{\phi} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = 0. \tag{3.9a}$$

This is satisfied by a solution of the corresponding Euler differential equation:

$$\frac{d^2y}{dx^2} = \frac{1}{2\phi}\left(\frac{\partial\phi}{\partial y} - \frac{dy}{dx}\frac{\partial\phi}{\partial x}\right)\left(1 + \left(\frac{dy}{dx}\right)^2\right), \tag{3.10}$$

which is directly derivable from the variation principle. Where the numerical values of the potential as a function of x and y are known, it is possible to perform a point-by-point integration of this equation—e.g., by the method of differences—and thus determine the trajectories of electrons in this field. Although extremely laborious, this is probably the most accurate method of obtaining electron paths.

3.5. Graphical Trajectory Determination. There are graphical methods for plotting electron paths on an equipotential map which are easy, rapid, and sufficiently accurate for most practical problems. Two of these are of sufficient importance as practical design tools to be worth discussing in detail.

The first is a circle method. Referring to Fig. 3.5, let it be assumed that an electron is moving in a potential field with the velocity indicated by the vector v_0. The magnitude of this velocity vector is:

$$v_0 = \sqrt{2\frac{e}{m}\phi_1}. \tag{3.11}$$

The electric field, E, is normal to the equipotential ϕ_1 and has a value approximately equal to

$$\frac{\phi_1 - \phi_2}{d} = E.$$

One component of the field lies along the direction of motion; the other, E_r, is at right angles to this direction. The latter exerts a radial force on the electron equal to eE_r, giving rise to a centripetal acceleration:

$$\frac{v^2}{r} = \frac{eE_r}{m}.$$

Solving for r and eliminating v with the aid of Eq. 3.11 this becomes:

$$r = 2\frac{\phi}{E_r}. \tag{3.12}$$

Accordingly, the path of the electron coincides approximately with the arc of a circle of radius r tangent to the vector v_0. Actually, this arc

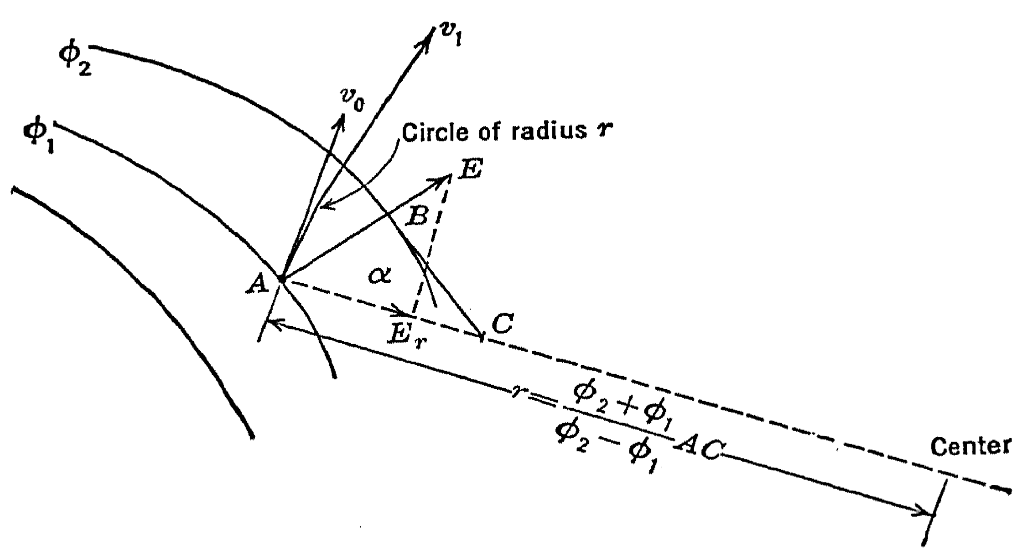


FIG. 3.5.—Circle Method of Graphical Ray Tracing.

should be infinitesimal in length, but since the equipotentials are close together, it may for this approximation be extended to meet the next equipotential, ϕ_2. At ϕ_2 the velocity vector will be tangent to the arc and will have a magnitude:

$$v_1 = \sqrt{2\frac{e}{m}\phi_2}.$$

By repeating the procedure at the successive equipotentials, the path of the electron can be mapped.

The radius and center of the arc can be found graphically, thus avoiding the calculation indicated in Eq. 3.12. First, the approximate direction of the field vector, E, is obtained by dropping a perpendicular from A, the intersection of the path with ϕ_1, onto the equipotential, ϕ_2, cutting

it at B. At right angles to the line AB a line is extended until it meets the normal to the velocity vector at C. It is evident that E_r must lie along the normal to the velocity vector and that the center of the arc must also be located on this line. If the angle between E and E_r is α, then:

$$E_r = E \cos \alpha = \frac{\phi_2 - \phi_1}{AB} \cos \alpha$$

$$r = 2\frac{\phi_2}{\phi_2 - \phi_1} \frac{AB}{\cos \alpha} = 2\frac{\phi_2}{\phi_2 - \phi_1} AC. \tag{3.13}$$

Somewhat greater accuracy is obtained if ϕ_2 in the numerator of Eq. 3.13 is replaced by the mean potential, giving:

$$r = \frac{\phi_2 + \phi_1}{\phi_2 - \phi_1} AC. \tag{3.13a}$$

With the aid of this construction, path-plotting can be carried out rapidly and accurately.

When the path in question starts from a surface of zero potential it is convenient to make use of the fact that it issues normal to the surface and has an initial radius of curvature three times as large as that of the electrostatic field lines in the neighborhood of its point of origin.

The second graphical procedure is the parabola method. This method is advantageous where the curvature of the path is small, which makes the radii awkwardly large when the circle method is used. Theoretically, if the process is carried to the limit and the separation between the equipotential lines made to approach zero, either method gives the true path. In all practical cases tested, the accuracy of the two methods is about the same.

The parabola method is based upon the fact that an electron moving in a uniform field having a velocity component at right angles to the field follows a parabolic path. It utilizes the geometric principle illustrated in Fig. 3.6a, namely, that the tangent to a parabola at a point at an axial distance x from its vertex meets the axis at point C, located at $-x$, or at an axial distance $2x$ from the point of tangency.

Again referring to Fig. 3.6a, let it be assumed that an electron at point A is moving with a velocity v, as indicated by the vector v, and that its motion is due to a uniform field. It is possible to determine the parabola giving its motion as follows: The component of velocity due to the electron having fallen through the potential field to point A is $v \cos \theta$, where θ is the angle between the velocity and field vectors. The difference of

potential ϕ^* between point A and the vertex of the parabola will, therefore, be:

$$\phi^* = \frac{1}{2}\frac{m}{e}v^2 \cos^2 \theta. \tag{3.14}$$

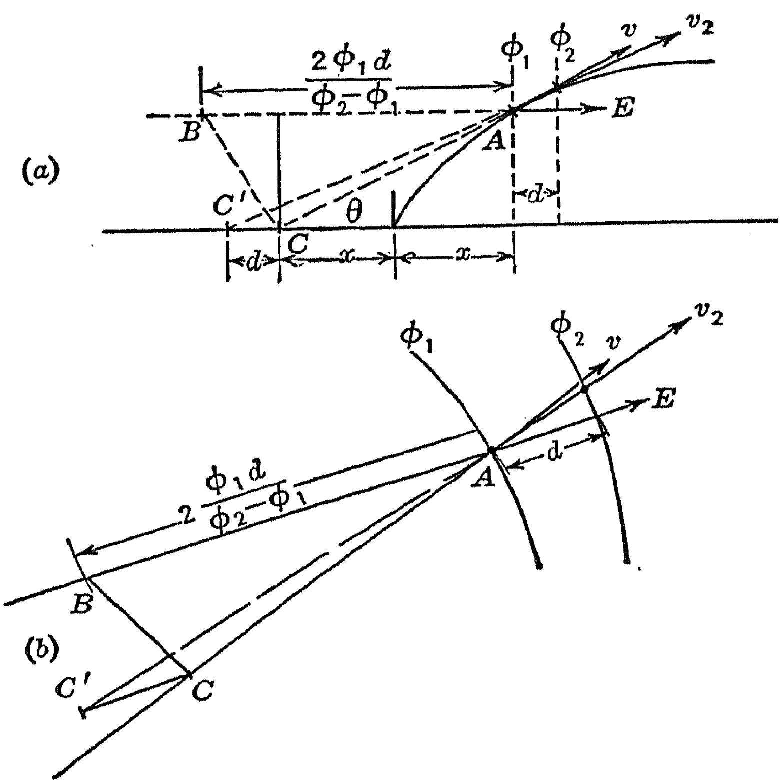


FIG. 3.6.—Parabola Method of Ray Tracing.

If two equipotentials, ϕ_1 and ϕ_2, are separated by a distance d, the field, E, can be expressed as $(\phi_2 - \phi_1)/d$. The distance between the vertex of the parabola and the point A can consequently be written as:

$$x = \frac{\frac{1}{2}\frac{m}{e}v^2 d \cos^2 \theta}{\phi_2 - \phi_1}$$

$$= \frac{\phi_1}{\phi_2 - \phi_1} d \cos^2 \theta. \tag{3.15}$$

Eq. 3.15 indicates a simple construction which will locate the point C, as shown in Fig. 3.6a. The field vector, E, is extended back a distance $2\phi_1 d/(\phi_2 - \phi_1)$ to B. From B a perpendicular is dropped to an extension of the velocity vector. This perpendicular will cut this vector at point C. Thus, this point can be determined from a knowledge of the vectors v and E. Further, if a line parallel to E is drawn through C, and extended back a distance d to C', this new point must lie on the velocity vector

for the point on the parabola where it intersects the equipotential ϕ_2. If a line is drawn through C' and A, it closely represents the path between the equipotentials ϕ_1 and ϕ_2.

To apply this construction to a general two-dimensional potential field, the procedure is as follows: Referring to Fig. 3.6*b*, the electron is at A, moving with a velocity and direction given by v. From point A, a line E representing the field direction is drawn normal to the equipotential, ϕ_1, and extends a distance d to ϕ_2. This line is drawn back from A a distance $2\phi_1 d/(\phi_2 - \phi_1)$ to point B. A perpendicular is then dropped from B onto the prolongation of v, locating point C. From C a line parallel to AB is drawn back a distance d, locating the point C'. A line through C' and A locates the position of the electron on the equipotential ϕ_2, and gives the direction of its velocity vector.

As the curvature of the path decreases, this method becomes increasingly accurate. It, therefore, is useful for determining the straighter portions of an electron trajectory, where, as has already been mentioned, the circle method becomes awkward because of the long radii involved.

The two plotting methods just discussed can be applied in any problem where the motion of the electron is confined to a plane; thus, it applies to any electrode configuration whose potential can be correctly mapped in a plotting tank of the type described.

3.6. The Rubber Model. By far the most convenient method of obtaining electron paths is by means of the rubber model. This model can be used in all problems where the potential can be expressed as a function of two rectangular coordinates, and where the electron path is confined to the plane of these coordinates. The accuracy which can be obtained is quite high, but not quite equal to that of a path carefully plotted by the graphical methods described.

A rubber membrane, stretched over a frame, is pressed down over a model of the electrode system which is made in such a way that its plan view corresponds to the geometrical configuration of the electrodes in the $x - y$ plane while the height is proportional to the negative voltage on each electrode. The rubber is then no longer flat, nor does it follow the surfaces of the model electrodes, but rather it stretches over them in a series of mountains and valleys, touching only the top of every electrode. If care is taken that the membrane is in contact with the full length of the top edges of all the electrodes the contours of its surface are found to correspond to an equipotential map of the electrode system.

A small sphere is placed at a point corresponding to the electron source, and allowed to roll on the rubber. The horizontal projection of its path then is a map of an electron trajectory in the electrode system under investigation.

The proof that the path of the sphere correctly represents an electron trajectory is divided into two parts. First, it is necessary to show that the height of every point on the rubber model is proportional to the potential existing in the electrode system. Second, it must be proved that a sphere rolling on such a surface follows a trajectory which represents the electron motion.

In order to show that the height z of the rubber surface represents the potential distribution, it is necessary to show that the surface obeys the differential equation:

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0. \tag{3.16}$$

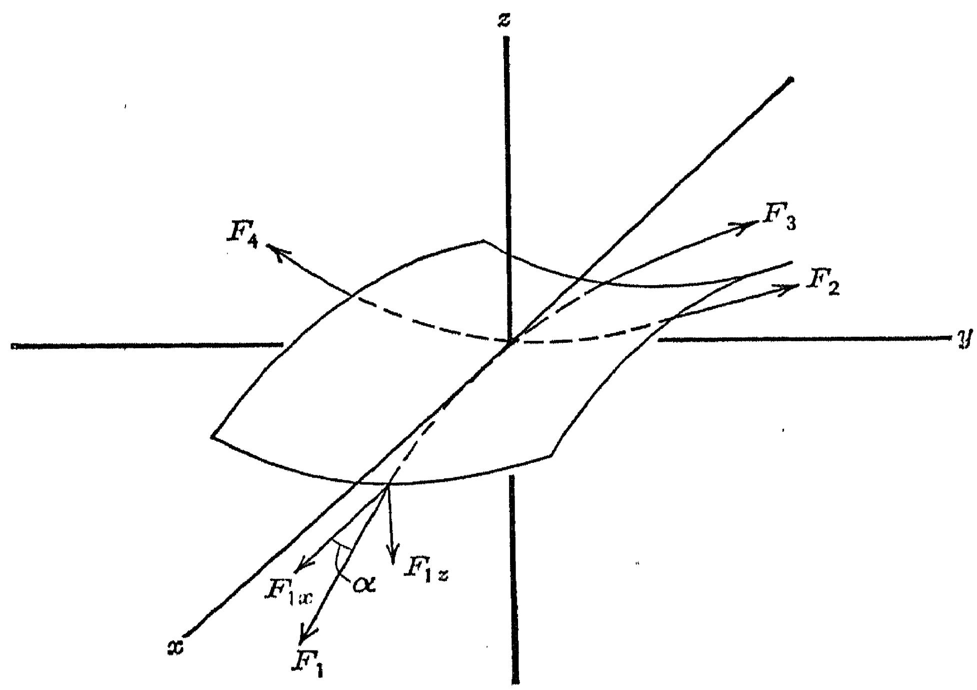


FIG. 3.7.—Forces Acting on an Element of a Stretched Membrane.

This may readily be demonstrated if two restrictions, which are more or less fulfilled in practice, are imposed on the rubber surface. These are: (1) that the slope of the surface be everywhere small, (2) that the tension of the deformed rubber be uniform over the surface.

The most straightforward proof applies the principle of minimum energy. Since the energy in any region is proportional to the area, the area integral of the surface must be a minimum. Transforming this minimized integral into the Euler form leads to the differential equation required.

The physical significance of the shape assumed by the surface is made more apparent by the following less rigorous demonstration.

Fig. 3.7 shows an element of surface area, ds, together with the forces acting on it. It is obvious that the vector sum of these forces must be

zero since the element is in equilibrium. In order to set up the conditions of equilibrium, the four forces F_1, F_2, F_3, and F_4 must each be resolved into their components along the coordinates. Considering first F_1, it is evident from the figure that:

$$F_{1x} = F_1 \cos \alpha = F_1 \frac{dx}{dx} \frac{1}{\sqrt{1 + \left(\frac{dz}{dx}\right)^2}}, \tag{3.17a}$$

$$F_{1z} = F_1 \sin \alpha = F_1 \frac{dz}{dx} \frac{1}{\sqrt{1 + \left(\frac{dz}{dx}\right)^2}}, \tag{3.17b}$$

where dz, dx, and α are as indicated. By the first restriction, dz/dx is small, so that

$$\frac{1}{\sqrt{1 + \left(\frac{dz}{dx}\right)^2}} \cong 1,$$

and the components become:

$$F_{1x} \cong F_1,$$
$$F_{1z} \cong F_1 \left(\frac{dz}{dx}\right)_1.$$

Similarly:

$$F_{3x} \cong F_3,$$
$$F_{3z} \cong F_3 \left(\frac{dz}{dx}\right)_3.$$

The two x components must be equal in magnitude and opposite in sign, hence:

$$F_{3x} \cong - F_{1x},$$
$$F_3 \cong - F_1,$$

and

$$F_{3z} = - F_1 \left(\frac{dz}{dx}\right)_3.$$

Summing the z components, the upward force due to F_1 and F_3 is:

$$F_{1z} + F_{3z} = F_1 \left(\left(\frac{dz}{dx}\right)_1 - \left(\frac{dz}{dx}\right)_3\right) = F_1 \frac{\partial^2 z}{\partial x^2} \Delta x, \tag{3.18}$$

where Δx is the length of the element in the x direction.

Applying the second restriction, that the force F_1 must equal the tension δ of the membrane times the width Δy of the element in the y direction, Eq. 3.18 becomes:

$$F_{1z} + F_{3z} = \delta\frac{\partial^2 z}{\partial x^2}\Delta x\Delta y. \qquad (3.18a)$$

In like manner, the vertical components of the other two forces are:

$$F_{2z} + F_{4z} = \delta\frac{\partial^2 z}{\partial y^2}\Delta y\Delta x.$$

Finally, since the sum of all the z force components must be zero, and δ, Δx, and Δy are not zero, the equation:

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0$$

must be true, proving that z satisfies the Laplace equation with the boundary conditions determined by the electrode heights.

It may be mentioned at this point that the slope of the rubber surface is everywhere proportional to the field strength, and that the force exerted on any electrode by the rubber is proportional to the capacity of that electrode.

The next problem is to show that a body moving under the action of gravity on the rubber surface moves along a path which corresponds to the electron trajectory. For simplicity, let it be assumed that the body in question slides on the surface, and that its friction is negligible.

By the principle of least action, the action integral must be stationary, or:

$$\delta\int_A^B mv\,ds = 0. \qquad (3.8)$$

Since the system is conservative and, hence, the sum of the kinetic and potential energies must remain constant, the momentum mv can be found as follows:

$$\text{K.E.} + \text{P.E.} = \text{const.},$$

$$\frac{1}{2}mv^2 = -mgz,$$

$$mv = \text{const.}\sqrt{z}.$$

By substitution Eq. 3.8 becomes:

$$\delta\int_A^B \sqrt{z}\,ds = \delta\int_A^B \sqrt{z}\sqrt{1 + \left(\frac{dy}{dx}\right)^2 + \left(\frac{dz}{dx}\right)^2}\,dx = 0. \qquad (3.19)$$

It has already been assumed that the slope of the rubber is small so that $(dz/dx)^2$ can be neglected compared with unity; therefore the final form of the action integral is:

$$\delta \int_A^B \sqrt{z} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = 0. \tag{3.19a}$$

Except for z replacing ϕ, Eq. 3.19a is seen to be identical with Eq. 3.9a. In the previous derivation it was shown that the height, z, of the rubber is proportional to the potential ϕ. Therefore, the path of the body sliding on the rubber is geometrically similar to the corresponding electron trajectory.

If, instead of the motion of a sliding body, that of a rolling sphere is considered, the conclusions are the same, provided that the assumption is made that the radius of curvature R of the sphere is small compared with the radius of curvature of the rubber. This is shown by deriving the total kinetic energy as follows:

Let $d\alpha$ be a small rotation of the sphere. Then the displacement, ds, of the center of the mass is given by:

$$ds = R\,d\alpha.$$

The angular velocity in terms of the linear velocity is thus:

$$\omega = \frac{d\alpha}{dt} = \frac{1}{R}\frac{ds}{dt} = \frac{v}{R}.$$

Next, the sum of the rotational and translational kinetic energies is expressed as follows:

$$\begin{aligned} \text{K.E.} &= \frac{1}{2}(mv^2 + I\omega^2) = \frac{1}{2}\left(m + \frac{I}{R^2}\right)v^2 \\ &= \frac{1}{2}m^*v^2, \end{aligned}$$

where m^* is the effective mass.

As before, the total momentum can now be written as $\sqrt{z}$ times a constant. Hence, Eq. 3.19a, expressing the principle of least action, is unchanged, and the path of the rolling sphere indicates the desired path.

Certain assumptions were made in deriving the paths on the rubber model, and the question might well arise as to how closely these assumptions must be fulfilled in order not to introduce serious errors in the final results. As a result of a large number of tests on the model, the indications are that, even if the slopes become as great as 30° to 45°, and the tension in the rubber very far from uniform, the paths obtained will be

sufficiently accurate for all practical purposes. In fact, the presence of friction, which has been neglected in the above derivation (except the implied friction required to produce rolling), makes it advisable to use fairly steep slopes.

A practical form of the rubber model is shown in Fig. 3.8. Ordinary surgical rubber is stretched over a square frame, which is about 3 feet on a side. Usually the electrode models are made from soft metal strips, either lead or aluminum, which are cut to the correct height to represent

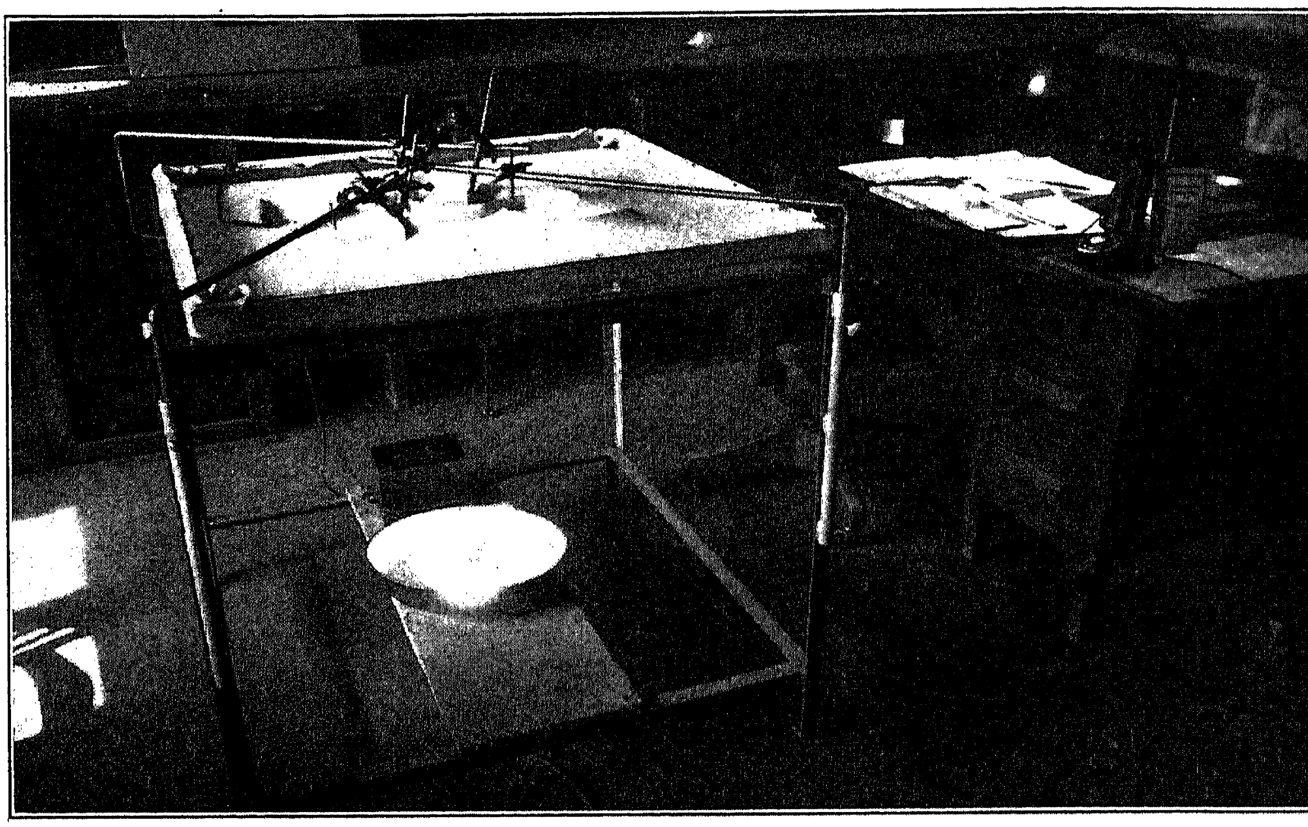


FIG. 3.8.—Rubber Model for the Study of Electron Paths.

the potential, and bent to conform with the electrode shape. The table supporting the electrodes is built of welded angle iron and has a plate-glass top. The glass top permits illumination from below, which greatly facilitates the placing of the electrodes. Since it is often necessary to press the rubber down to make it come in contact with the more positive electrodes, the table is equipped with movable side arms to which can be clamped top electrode models.

It has been found convenient to use 3/16-inch steel ball bearings for the spheres. These have an advantage over glass spheres in that they can be held at the top of the cathode electrode with a small electromagnet and released when desired, without any danger of deflecting their course, merely by cutting off the current to the magnet.

Where a permanent record of the path is desired, a time exposure for

the duration of the motion of the sphere can be made. In this case it is better to use black rubber for the model, and to illuminate it from above. Furthermore, with a pulsating light source, such as a 60-cycle cold arc,

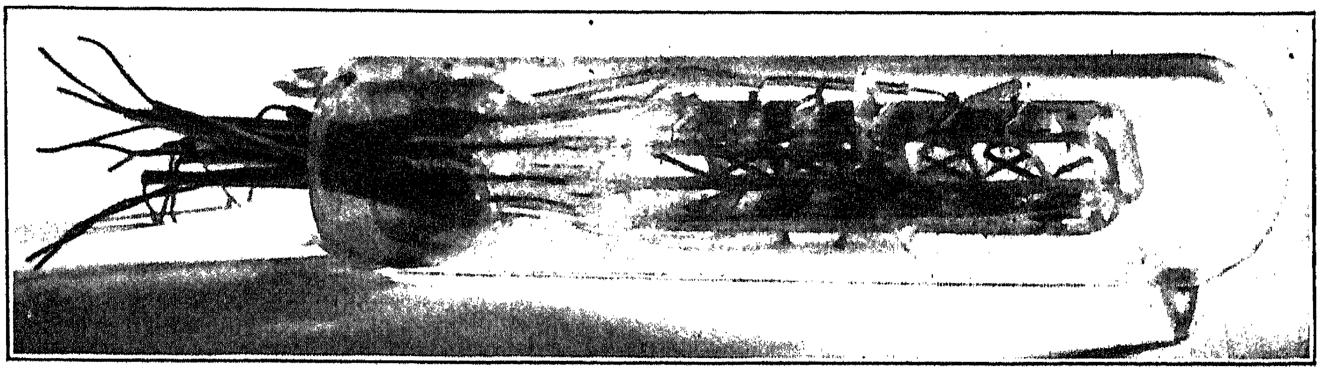


FIG. 3.9.—An Experimental Electrostatic Secondary-Emission Multiplier.

the paths will appear as dotted lines. The spacing between the dots is a measure of the velocity of the electron.

Numerical values for the accuracy that can be attained with the rubber model are difficult to give, since the error depends upon the elec-

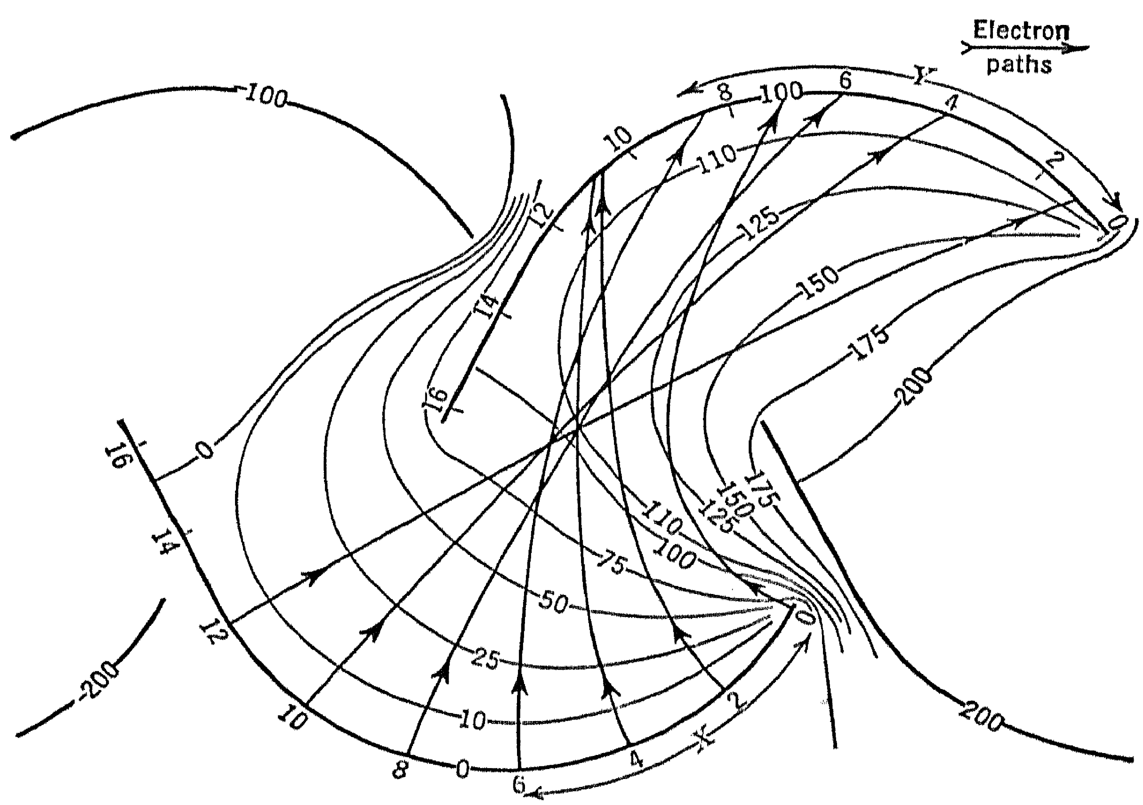


FIG. 3.10.—Electron Paths in a Multiplier of the Type Shown in Fig. 3.9 as Determined from Potential Plot.

tron path. Measurements of parabolic paths obtained on such a model were made by P. H. J. A. Kleynen, who found an error of 1 per cent in the height of the parabola, and one of 7 per cent in the separation be-

tween the arms of the parabola when the sphere returned to its initial potential.

Results obtained with the model when used in connection with the design of the electrostatic multiplier shown in Figs. 3.9 and 3.10 indicate its excellence. Plots were made of the initial and terminal points of a number of electrons, as indicated by the model, and again with an actual

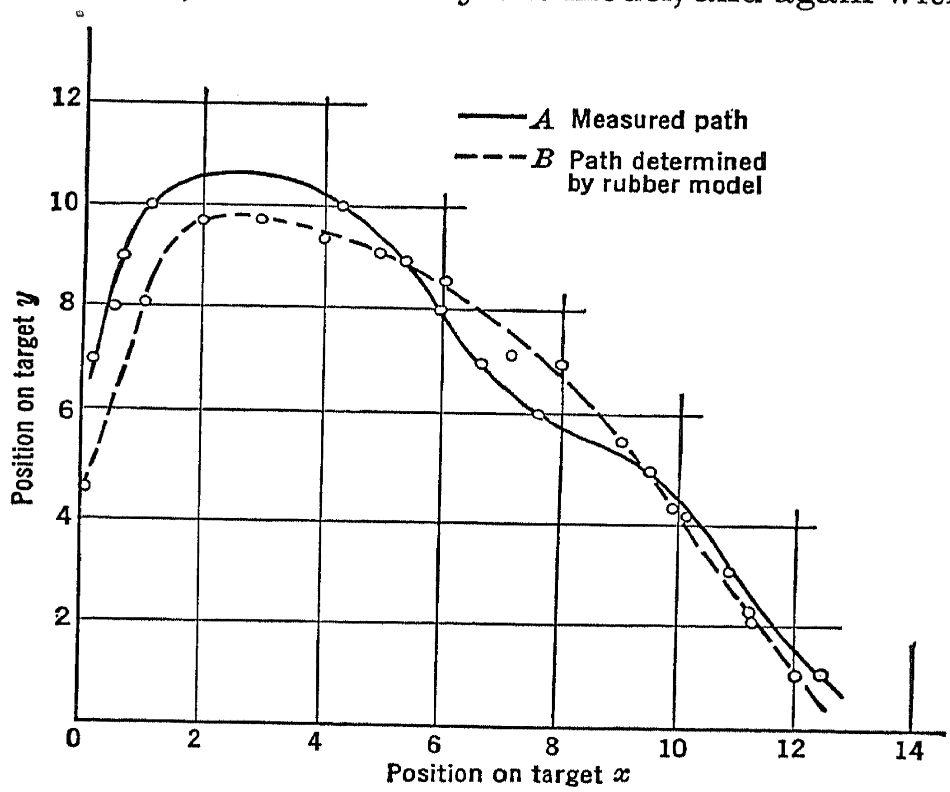


FIG. 3.11.—Curves Showing Termini of Electron Trajectories in the Multiplier as Determined by Direct Measurement and from a Rubber Model.

electron tube. Two curves of this type are reproduced in Fig. 3.11, showing fairly close agreement between the two methods of measurement.

REFERENCES

1. J. H. JEANS, "Electricity and Magnetism," Cambridge University Press, 1927.
2. M. MASON and W. WEAVER, "The Electromagnetic Field," University of Chicago Press, 1929.
3. O. D. KELLOGG, "Foundations of Potential Theory," Julius Springer, Berlin, 1929.
4. D. GABOR, "Mechanical Tracer for Electron Trajectories," *Nature*, Vol. 139, p. 373, Feb. 27, 1937.
5. H. SALINGER, "Tracing Electron Paths in Electric Fields," *Electronics*, Vol. 10, pp. 50–54, October, 1937.
6. P. H. J. A. KLEYNEN, "The Motion of an Electron in a Two-Dimensional Electrostatic Field," *Philips Tech. Rev.*, Vol. 2, pp. 321–352, 1937.
7. J. R. PIERCE, "Electron Multiplier Design," *Bell Tel. Record*, Vol. 16, pp. 305–309, 1938.
8. V. K. ZWORYKIN and J. A. RAJCHMAN, "The Electrostatic Electron Multiplier," *Proc. I. R. E.*, Vol. 27, pp. 558–566, Sept., 1939.

CHAPTER 4

ELECTRON OPTICS (*Continued*)

The electron-optical systems to be considered in this chapter are those with cylindrically symmetrical field-producing elements. All electron lenses are based on configurations of this type. Conversely it may be stated quite generally that any varying electrostatic or magnetic field which has cylindrical symmetry is capable of forming a first-order image, either real or virtual.

The practical importance of electron lenses is only now becoming apparent, although they have for some time attracted considerable attention in the realm of pure science.

Perhaps the first practical application of the electron lens was in the electron gun of cathode-ray oscilloscopes. This has been greatly improved in the past few years and is now used for television purposes in the Kinescope and the Iconoscope. Details of the electron gun, although it embodies principles discussed here, will be reserved for a later chapter.

Another early use of the electron lens is found in the electron microscope. There are a great many forms of this microscope, all based essentially on the same principle. Fig. 4.1 illustrates a high-magnification instrument utilizing magnetic electron lenses. Not only is the electron microscope an important aid in the study of cathodes, secondary emitters, and metal surfaces, but, in addition, it has recently been adapted to biological work, permitting higher useful magnifications than can be obtained by optical means.

A third application which should be mentioned is its use in connection with the image tube. The image tube is of importance because it can be combined with the Iconoscope to make a television pickup tube which is many times more sensitive than the normal Iconoscope. Here again there are a number of possible forms, using electrostatic, magnetic, or combined electrostatic and magnetic lenses.

In order to make intelligent use of electron lenses it is necessary to have a rather complete understanding of the process of image formation. The theory of the formation of images in electron optics may be regarded as an extension of that applying to light optics. It is, therefore, not out of place to review briefly the elements of ordinary optics.

4.1. Optical Principles. When a ray of light passes through a boundary between two media in which the velocity of light differs, the ray is bent by a process known as refraction. The law governing this refraction is the well-known Snell's law:

$$n \sin \beta = n' \sin \beta', \tag{4.1}$$

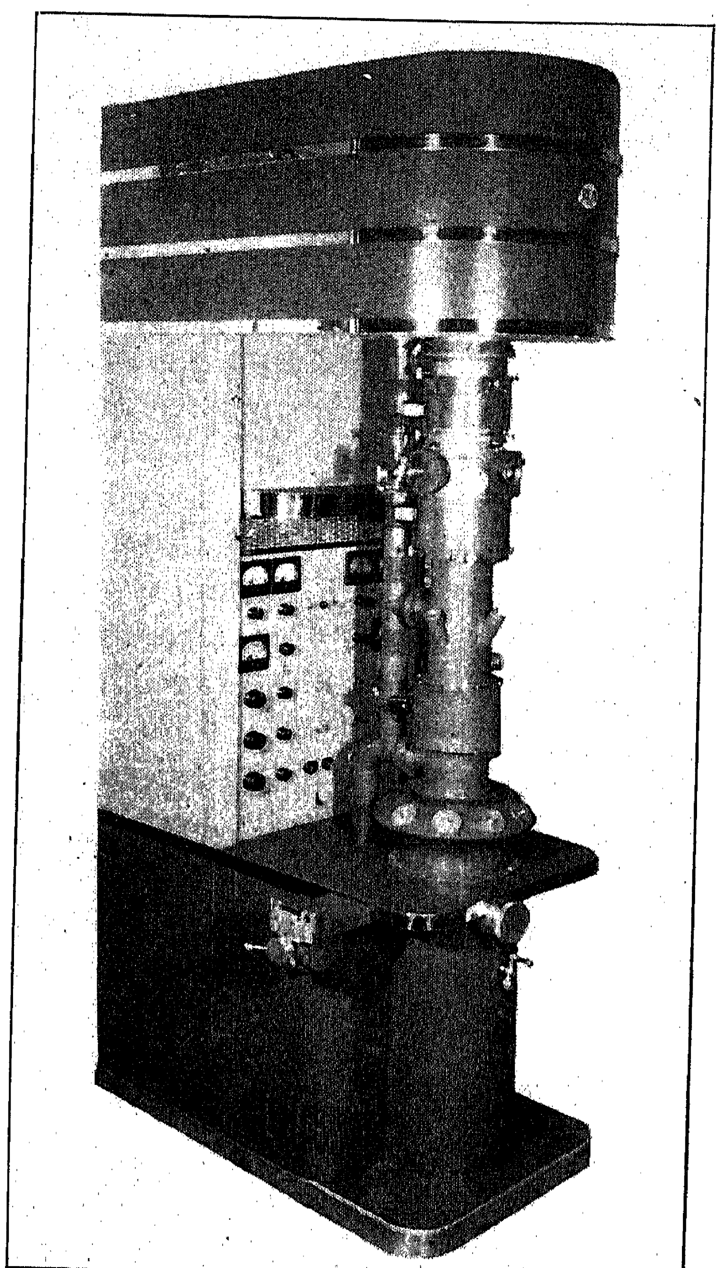


FIG. 4.1.—R.C.A. Electron Microscope.

where β and β' are the angles which the incident and refracted ray make with the normal to the boundary between the media having refractive indices of n and n', respectively.

If the boundary separating the two media is a section of a spherical surface, a lens will be formed. Such a surface can be shown to have

image-forming properties. By this is meant that, if the light rays from a small object enter the spherical refracting surface, the rays from any point will be bent in such a way that they converge on, or appear to diverge from, a second point known as its image point. Furthermore, the image points will be ordered in the same way as the emitting or object points, so that an image is formed of the original object. Where the rays travel so that they actually converge on the image points, the image

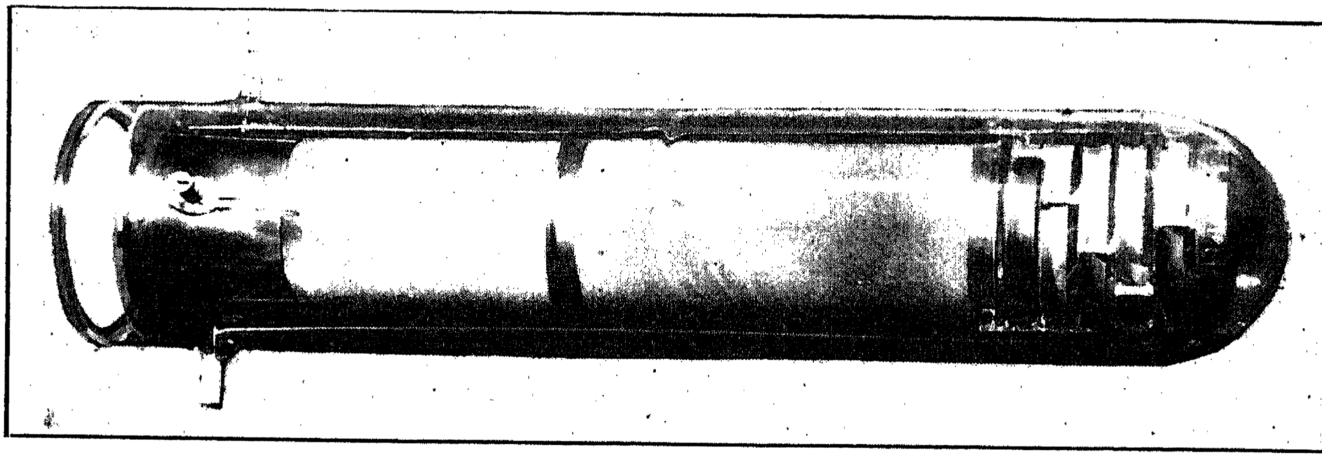


Fig. 4.2.—An Electrostatic Image Tube.

formed is said to be real; if it is necessary to extend the rays backward to the point from which they appear to diverge, the image is virtual.

To prove the existence of the image-forming property of these spherical surfaces, it is necessary to show, first, that the rays from an object point converge on an image point, and second, that the ordering of object and image points is similar. The carrying out of this demonstration requires the imposing of two restrictions: namely, that the object

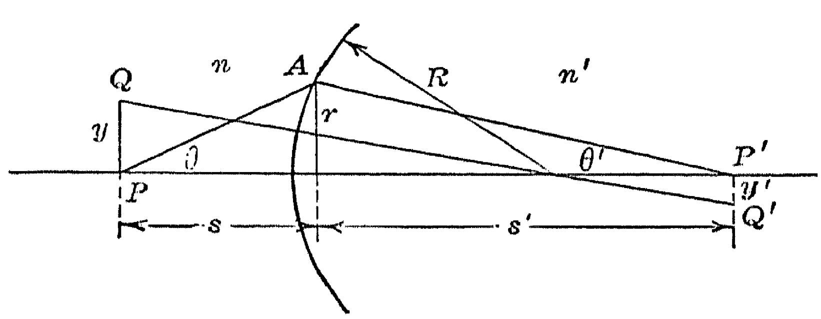


Fig. 4.3.—Refraction at a Curved Surface.

tion requires the imposing of two restrictions: namely, that the object be small, and that the rays make very small angles with the normal from the object point to the surface. Rays which meet these requirements are known as paraxial rays, and the image theory based on these restrictions is called the first-order theory, or Gaussian dioptrics.

Referring to Fig. 4.3, P is the object point at a distance s from the

spherical boundary of radius R. The ray PA emitted at an angle θ is refracted so that it reaches the axis at s', making an angle θ' with the axis. From the geometry of the figure it is obvious that the angles of incidence and refraction are:

$$\beta = \theta + \frac{r}{R},$$

$$\beta' = -\theta' + \frac{r}{R}.$$

Since, under the restrictions imposed, the angles of incidence and refraction are small, the sines appearing in Snell's law may be replaced by the angles themselves. Eq. 4.1 thus becomes:

$$n\beta = n'\beta' \tag{4.2}$$

and, substituting in Eq. 4.2, it follows that

$$n\theta + \frac{rn}{R} = -n'\theta' + \frac{rn'}{R}$$

or

$$n\theta + n'\theta' = \frac{r}{R}(n' - n). \tag{4.3}$$

However, $\theta = r/s$ and $\theta' = r/s'$, so that Eq. 4.3 can be written as follows:

$$\frac{n}{s} + \frac{n'}{s'} = \frac{n' - n}{R}. \tag{4.4}$$

Since neither r nor θ appears in this expression, the equation proves that any ray leaving P must converge on the point P' at a distance s' from the surface. The point P' is therefore the image point of the object point P.

Considering next a particular ray QQ', which is normal to the refracting surface, it is evident from the similar triangles formed that

$$\frac{y'}{y} = \frac{(s' - R)}{R + s}.$$

Combined with Eq. 4.4 this becomes:

$$\frac{y'}{y} = \frac{ns'}{n's}. \tag{4.5}$$

Since y'/y is the ratio of the separation of the two object points PQ, and the two image points $P'Q'$, it is the lateral magnification m of the image. This magnification is independent of which point on the image is chosen;

therefore, the ordering of the image and object points will be similar, and the second condition for image formation is fulfilled.

Finally, substituting the relation $s'/s = \theta/\theta'$ into Eq. 4.5, the following equation is obtained:

$$ny\theta = n'y'\theta'. \tag{4.6}$$

This equation expresses the law of Lagrange and is valid for the first-order image formation by any number of refracting surfaces.

Referring back to Fig. 4.3, it is evident that it is possible to conceive of a plane surface which would bend the rays in exactly the same way as the spherical surface. As can be seen from the figure, the amount of bending required to produce an image would be:

$$\alpha = kr, \tag{4.7}$$

where α is the angle between the incident and refracted ray, and r the radial distance from the axis at which the ray meets the surface.

4.2. The Thin Lens. The formulas just derived are sufficient to determine the size and position of the first-order image formed by any number or type of curved refracting surfaces. However, it is often sim-

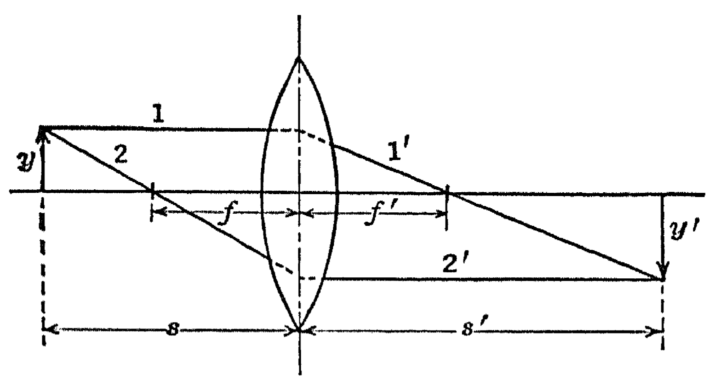


FIG. 4.4.—Properties of a Thin Lens.

pler to express the properties of the lens in another way. When dealing with the image-forming properties of a thin lens, such as is illustrated in Fig. 4.4, i.e., a refractive medium bounded by two spherical surfaces having radii of curvature large compared to the separation between them, it is most convenient to make use of the derived concept of focal length.

Since, for this lens, the index of refraction in the image and object space is the same, it follows from the successive application of Eq. 4.4 that

$$\frac{1}{s} + \frac{1}{s'} = \text{const.}$$

If, as for ray 1, s is made equal to infinity, this becomes:

$$\frac{1}{\infty} + \frac{1}{f'} = \text{const.},$$

or, similarly, for ray 2′, where s' is infinite:

$$\frac{1}{f} + \frac{1}{\infty} = \text{const.}$$

When combined, these equations give:

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} = \frac{1}{f'}, \tag{4.8}$$

which is the familiar equation for a thin lens where $f = f'$ is by definition the focal length. It will be noted that the focal length of the system represented by Eq. 4.7 is $1/k$. The magnification, as is evident from the geometry of the figures, will be in either case:

$$m = -\frac{s'}{s}. \tag{4.9}$$

4.3. The Thick Lens. The relations developed above can be applied with a fair degree of accuracy for all lenses whose focal length is long

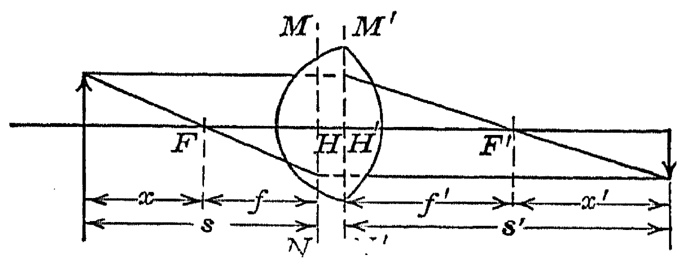


FIG. 4.5.—Image Formation by a Thick Lens.

compared to the thickness. However, if the thickness of the lens cannot be considered as negligible compared to the focal length, a more complicated set of relations is required. These apply to a single lens or to a system composed of several lens elements. Either system is spoken of as a "thick lens."

In order to determine the size and position of the Gaussian image formed by a thick lens, it is necessary to locate two reference planes, known as "principal" planes, and the two focal points.

The two principal planes are the conjugate planes for which the optical system has a positive magnification of unity. In other words, an object at one principal plane produces at the other a virtual (erect) image which is the same size as the object. The intersections of these planes with the axis of the system are known as the principal points. Fig. 4.5 shows a thick-lens optical system, the planes MN and $M'N'$ being the two principal planes, and H and H' the principal points. Any ray of light entering the lens from the object space parallel to the axis will be bent

in such a way that it crosses the axis at the point F'. The parallel ray, and the ray through the focal point, will, if extended, meet at the principal plane of the image space $M'N'$. Similarly, any ray through the point F in the object space will emerge from the lens system parallel to the axis. These two rays, when extended, meet in the first principal plane MN. The points F and F' are the first and second focal points, respectively. The two focal points, together with the two principal points, are known as the cardinal points of the lens system. Once these have been located, the first-order image of any object can readily be found. The distances f and f', between the focal points and their corresponding principal planes, are known as the first and second focal lengths.

From the geometry of Fig. 4.5 it is evident that the lateral magnification m of the system is given by:

$$m = -\frac{f}{x} = -\frac{x'}{f'}, \tag{4.10}$$

where x and x' are object and image distances from the focal points. From this it follows that:

$$xx' = ff'. \tag{4.11}$$

The magnification and position can, of course, be referred to the principal points instead of the focal points. If s and s' are the distances of the object and image from the principal planes, their values are $s = x + f$ and $s' = f' + x'$. When these are substituted in Eq. 4.11, the latter becomes:

$$(s - f)\,(s' - f') = ff', \tag{4.12}$$

$$\frac{f}{s} + \frac{f'}{s'} = 1,$$

which corresponds to Eq. 4.8 for a thin lens. In the same terms, the magnification is:

$$m = -\frac{f}{s - f} = -\frac{s' - f'}{f'}. \tag{4.13}$$

It can be shown that, if the medium in the object space has a refractive index n and that in the image space is n', the two focal lengths will be related as follows:

$$\frac{f}{f'} = \frac{n}{n'}. \tag{4.14}$$

As a consequence, if the media on the two sides of the system have the

same refractive index, the first and second focal lengths will be equal. A derivation of these laws dealing with a thick lens can be found in any elementary textbook on optics.

As was mentioned above, any combination of thin lenses may be represented by the four cardinal points of a thick lens. Two thin lenses of focal lengths f_1 and f_2 separated by a distance d will serve as an illustrative example. Such a system is shown in Fig. 4.6. In the following, all quantities referring to the first lens will have the subscript 1; those to the second lens, the subscript 2; those referring to the equivalent thick lens

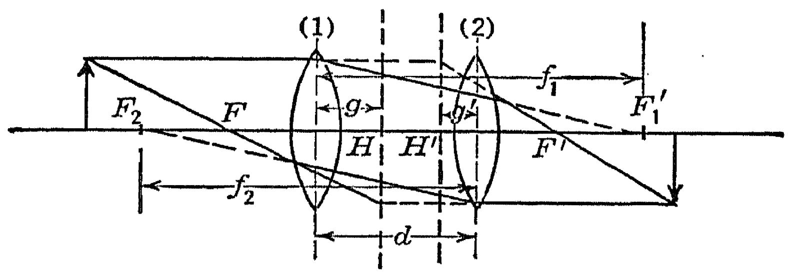


FIG. 4.6.—Cardinal Points of a Pair of Thin Lenses.

will have no subscript. An object at infinity, making $s_1 = s = \infty$, is imaged at F_1' by the first lens. This image is the object for the second lens, the object distance being:

$$s_2 = d - f_1.$$

The point at which the second lens images this virtual object will be the position of second focal point F' of the equivalent thick lens. The distance between F' and the second lens is, therefore, s_2'. Applying Eq. 4.12 it follows that:

$$\left.\begin{aligned} \frac{1}{d - f_1} + \frac{1}{s_2'} &= \frac{1}{f_2}, \\ s_2' &= \frac{f_1 f_2 - d f_2}{f_1 + f_2 - d}, \end{aligned}\right\} \tag{4.15}$$

which locates the second focal point. The first focal point can be located in a similar way by tracing a ray to an image point at infinity. The distance between the first lens and the first focal point will be found to be

$$\frac{f_1 f_2 - d f_1}{f_1 + f_2 - d}. \tag{4.15a}$$

Furthermore, from the geometry of the two sets of ray paths discussed, it can be shown that the distances g and g' between the first principal

point and first lens, and between the second principal point and second lens, are:

$$g = \frac{df_1}{f_1 + f_2 - d}, \tag{4.16}$$

$$g' = \frac{df_2}{f_1 + f_2 - d}. \tag{4.16a}$$

In this way, the four cardinal points of the equivalent thick lens can be located.

4.4. Index of Refraction in Electron Optics. From the foregoing it is evident that the concept of the index of refraction is important in optics.

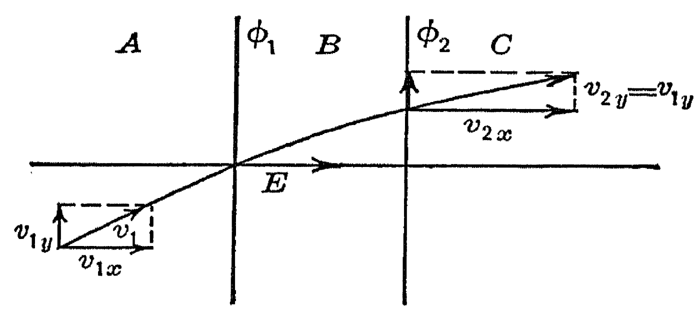


FIG. 4.7.—Refraction by Potential Double Layer.

In electron optics the potential, or rather the square root of the potential, plays the role of the index of refraction. The following simple example illustrates the similarity. Consider an electron moving from a region at potential ϕ_1 through a narrow transition region into a region at potential ϕ_2. Referring to Fig. 4.7, the regions A and C are field-free and at potentials ϕ_1 and ϕ_2, respectively, while in the intervening transition space B there exists a field of magnitude $(\phi_2 - \phi_1)/d$ normal to the boundary sheets. The boundaries separating the three regions are assumed to be conducting sheets which are transparent to electrons. Such sheets are, of course, purely fictitious and can only be approximated in practice; however, they are useful for illustrative purposes. The incident electron approaching the first sheet has a velocity

$$v_1 = \sqrt{2\frac{e}{m}\phi_1}.$$

If it makes an angle θ_1 with the surface normal, its velocity can be resolved into two components, which are:

$$v_{1x} = \sqrt{2\frac{e}{m}\phi_1} \cos \theta_1, \tag{4.17}$$

$$v_{1y} = \sqrt{2\frac{e}{m}\phi_1} \sin \theta_1, \tag{4.17a}$$

where x is normal to the boundary sheets and y is parallel to them. As the electron traverses the transition region, it is accelerated in the x direction by the field, while its y component of velocity remains unchanged. The total velocity of the emerging electron will, of course, be:

$$v_2 = \sqrt{2\frac{e}{m}\phi_2}. \tag{4.18}$$

From Eq. 4.17*a* its transverse velocity is:

$$v_{2y} = \sqrt{2\frac{e}{m}\phi_1} \sin \theta_1. \tag{4.19}$$

The angle of emergence θ_2 is determined by these two velocities and is given by:

$$\frac{v_{2y}}{v_2} = \sin \theta_2.$$

Finally, substituting Eqs. 4.18 and 4.19 into this equation, and transposing, it becomes:

$$\sqrt{\phi_1} \sin \theta_1 = \sqrt{\phi_2} \sin \theta_2. \tag{4.20}$$

Comparing this to Snell's law, it will be seen that $\sqrt{\phi}$ is the exact counterpart of the index of refraction n.

This similarity is quite general, as can be proved by comparing Fermat's principle of optics:

$$\delta \int_A^B n ds = 0$$

and the principle of least action for an electron:

$$\delta \int_A^B v ds = \text{const.}\ \delta \int_A^B \sqrt{\phi}\, ds = 0.$$

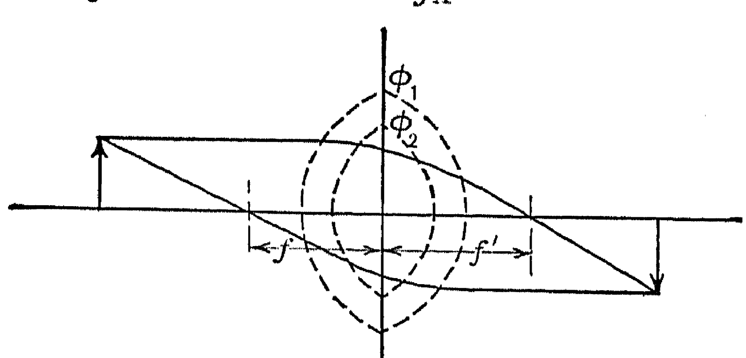


FIG. 4.8.—Double-Layer Lens.

4.5. Simple Double-Layer Lens. The simplest concept of an electron lens is probably that formed by two curved double layers of the type just described. The arrangement is illustrated in Fig. 4.8. The double

layer on the object side is assumed to have a radius of curvature R_1 and that on the image side a radius R_2. The potential of the inner surfaces is ϕ_2; that of the outer, ϕ_1.

This lens corresponds exactly to an optical thin lens with an index of refraction $n_2 = \sqrt{\phi_2}$ immersed in a medium which has an index of refraction $n_1 = \sqrt{\phi_1}$. Its focal length can be calculated by elementary optics and is given by the equation:

$$\frac{1}{f} = \left(\sqrt{\frac{\phi_2}{\phi_1}} - 1\right)\left(\frac{1}{R_1} + \frac{1}{R_2}\right).$$

An electron source, located at a distance s from this lens, emitting electrons with a velocity $\sqrt{2(e/m)\phi_1}$, will be imaged at a distance s'. These distances are related by the equation:

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f}.$$

4.6. Continuous Lenses. If the type of lens just described were the only kind that could be made, electron lenses would have little practical value. There exists a second, fundamentally different class, and upon this the importance of electron lenses rests. The operation of these lenses is based upon the following fact: Whenever there exists a region in which there is a varying electric field having cylindrical symmetry, this region will have properties analogous to those of an optical lens system; that is, it will be capable of forming a real or virtual image of an emitting source.

It is evident that a lens formed in this way is very different from the familiar glass lens treated in conventional optics. Instead of sharp boundaries between media of different index of refraction, the index varies continuously, both along the axis and radially. This being so, it has been found necessary to apply different mathematical methods in calculating the lens properties resulting from any specific field. These properties, once they have been found, can be represented by the same four cardinal points which describe an optical thick lens, namely, by two focal points and two principal points.

The simplest lens having a continuously variable index of refraction is that formed by an axially symmetric transition region between two constant fields of different magnitude. Physically, such a region is approximated by that in an aperture having different field strengths on the two sides. This is illustrated in Fig. 4.9a.

In order to obtain a physical picture of the action of this lens it is convenient to make use of the concept of lines of electrostatic force, or field

lines, whose direction at any point is that of the field and whose density is proportional to the field strength. A small volume element in the transition region of the lens is shown in Fig. 4.9*b*. In this region the Laplace equation is obeyed, and, as was pointed out in the preceding chapter, this requires that the number of field lines which enter the volume element must equal that which leave. It is evident that, since the density of lines is different on the two ends, lines must enter or leave through the top of the volume. Referring to Fig. 4.9*b* it will be seen that this condition can be fulfilled only if the field lines are curved. Accordingly, the field lines are not parallel to the axis, and the field can be resolved into radial and axial components. The radial component of the field will deflect any electron passing through the transition region toward (or away from) the axis. This bending action increases with radial distance from the axis, as is required for image formation.

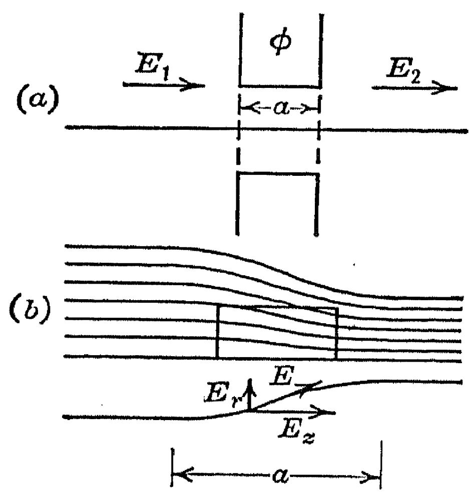


Fig. 4.9.—Idealized Field Distribution in Simple Aperture Lens.

The quantitative properties of the lens can be determined as follows*: First, by applying the Laplace equation the radial component of the field is obtained. From the radial field component the change in radial momentum of an electron in passing through this region is calculated. This leads to the value of the change in angle of the electron trajectory and, hence, directly to the focal length of the system.

Since there are no charges, the divergence as well as the curl of the field are zero. In cylindrical coordinates these can be expressed as follows:

$$\frac{\partial E_z}{\partial z} + \frac{1}{r}\frac{\partial}{\partial r}(rE_r) = 0, \tag{4.21}$$

and:

$$\frac{\partial}{\partial z}(rE_r) - r\frac{\partial E_z}{\partial r} = 0. \tag{4.22}$$

The total differential of the radial field component is:

$$d(rE_r) = \frac{\partial}{\partial z}(rE_r)dz + \frac{\partial}{\partial r}(rE_r)dr.$$

Substituting from Eqs. 4.21 and 4.22 this becomes:

$$d(rE_r) = r\frac{\partial E_z}{\partial r}dz - r\frac{\partial E_z}{\partial z}dr, \tag{4.23}$$

* See Bedford reference 13.

which gives the radial component in terms of the axial field. To determine the first-order-image properties of the lens only the field near the axis need be considered. In the vicinity of the axis both E_z and $\partial E_z/\partial z$ are to the desired approximation independent of r. Therefore, Eq. 4.23 may be written:

$$d(rE_r) = -r\frac{\partial E_z}{\partial z}dr,$$

and integrated directly to give:

$$E_r = -\frac{r}{2}\frac{\partial E_z}{\partial z}. \tag{4.24}$$

The change in radial momentum of an electron subjected to this field for a time equal to that required to traverse the transition region, whose width is a, is:

$$\begin{aligned} \Delta mv_r &= -\int eE_r dt \\ &= \int_0^a \frac{er}{2}\frac{\partial E_z}{\partial z}\frac{dz}{v_z}, \end{aligned} \tag{4.25}$$

since $dt = dz/v_z$. As a is small, v_z can be assumed to be constant over the transition region, and this equation can be integrated, giving:

$$\Delta mv_r = \frac{er}{2v_z}(E_2 - E_1).$$

The change in angle is, therefore:

$$\begin{aligned} \alpha &= \frac{\Delta mv_r}{mv_z} \\ &= \frac{er}{2v_z^2 m}(E_2 - E_1), \end{aligned}$$

or, replacing mv_z^2 by $2e\phi$, where ϕ is the potential of the lens:

$$\alpha = r\frac{E_2 - E_1}{4\phi}. \tag{4.26}$$

A comparison of Eq. 4.26 with Eq. 4.7 indicates that the lens will form a first-order image. Its focal length is:

$$f = \frac{4\phi}{E_1 - E_2}. \tag{4.27}$$

Although, in view of the restrictions assumed in the derivation, Eq. 4.27 is only a first approximation, it is nevertheless useful for estimating the behavior of apertures and other electron-optical systems. For example, a lens often encountered in practice consists of two apertures at

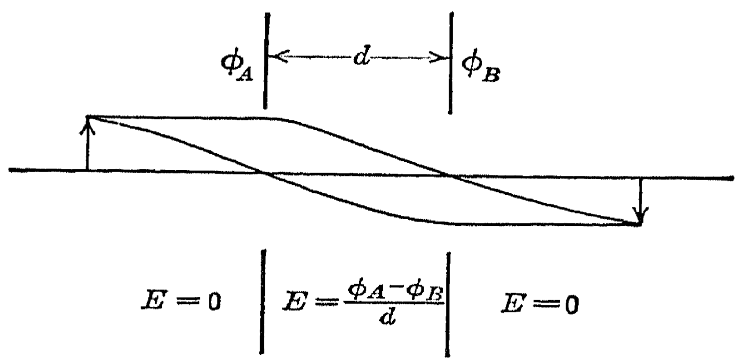


FIG. 4.10.—Ray Paths in Double-Aperture Lens (Schematic).

different potentials as illustrated in Fig. 4.10. The focal lengths of the apertures considered separately are:

$$f_A = \frac{4\phi_A d}{\phi_B - \phi_A},$$

$$f_B = \frac{-4\phi_B d}{\phi_B - \phi_A}.$$

This being combined into a "thick lens," due account being taken of the converging action of the field between the apertures and the difference in potential on the two sides of the lens, the focal lengths f and f' on the object and image sides of the system are:

$$\left.\begin{aligned} f &= \frac{8d}{3\left(\dfrac{\phi_B}{\phi_A} - 1\right)\left(1 - \sqrt{\dfrac{\phi_A}{\phi_B}}\right)}; \\ f' &= \frac{8d}{3\sqrt{\dfrac{\phi_A}{\phi_B}}\left(\dfrac{\phi_B}{\phi_A} - 1\right)\left(1 - \sqrt{\dfrac{\phi_A}{\phi_B}}\right)}. \end{aligned}\right\} \tag{4.28}$$

Again warning should be given that these results are only approximate, and are strictly applicable only when ϕ_A and ϕ_B are not too different, d is small compared with the focal length, and the aperture diameter small compared with d.

4.7. The Ray Equation. In most practical cases, the lens region, i.e., the region throughout which the field is varying, cannot be considered as small, as was assumed in the derivation given in the preceding section.

To determine the imaging properties of a more general electron lens it is necessary to make use of a differential equation of the ray path, which can be integrated over the region of varying potential. Proceeding as in the previous example, but considering as the transition region an ele-

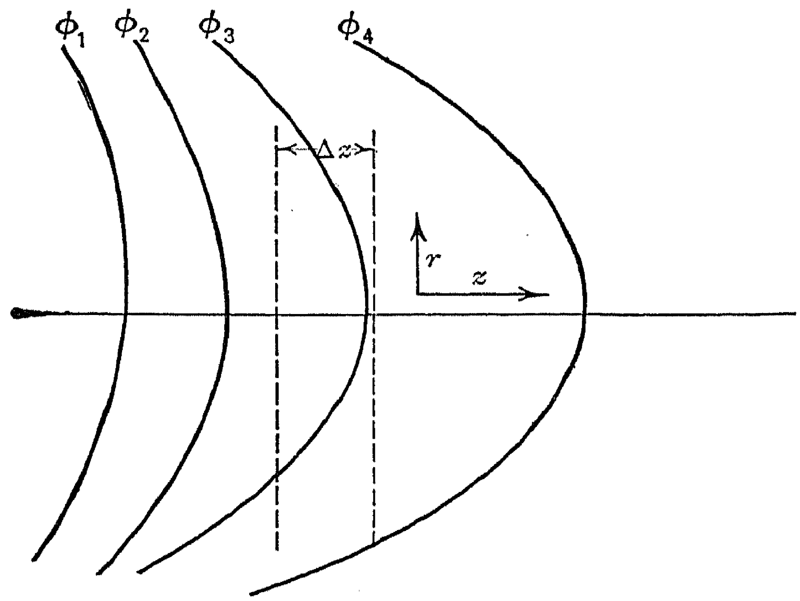


FIG. 4.11.—Elementary Section of the Potential Field of an Electron Lens.

mentary strip between z and $z + \Delta z$, which is part of an extended region of varying potential, the change in radial momentum is:

$$\Delta m v_r = \frac{re}{2} \frac{\partial E_z}{\partial z} \frac{\Delta z}{v_z};$$

making the substitutions $v_r = dr/dt$ and $\partial E_z/\partial z \cong - d^2\phi/dz^2$ this becomes:

$$\Delta \frac{dr}{dt} = \frac{re}{2m} \frac{d^2\phi}{dz^2} \frac{\Delta z}{v_z},$$

ϕ being the potential along the axis. Dividing through by Δz and letting the width Δz of the elementary strip approach zero, it follows that

$$\frac{d}{dz}\left(\frac{dr}{dt}\right) = - \frac{er}{2mv_z} \frac{d^2\phi}{dz^2}. \tag{4.29}$$

By a suitable rearrangement of terms and the substitution $v_z \cong v = \sqrt{(2e/m)\phi}$, Eq. 4.29 can be written as:

$$\frac{d^2r}{dz^2} = - \frac{1}{2\phi} \frac{d\phi}{dz} \frac{dr}{dz} - \frac{1}{4\phi} \frac{d^2\phi}{dz^2} r, \tag{4.30}$$

$$r'' = - \frac{r'\phi'}{2\phi} - \frac{r\phi''}{4\phi}.$$

This is the fundamental ray equation used in all cylindrically symmetric electron-optical problems involving only electrostatic fields. When solved it gives the radial distance of the electron ray from the axis at every point along the axis. Furthermore, its form assures the fulfilment of Gaussian, or first-order, image requirements.

4.8. Solution of the Ray Equation. When a solution of Eq. 4.30 is possible, the first-order-image properties of the electron lens system can be readily determined. The procedure is similar to that in optics. An electron ray parallel to the axis in object space is traced through the lens. Its intersection with the axis locates the second focal point. Extending a line tangent to the ray at the second focal point until it intersects the incident ray locates the position of the second principal plane. The first focal point and principal plane can be determined by similarly considering a ray parallel to the axis in image space. With the four cardinal points, the position and magnification of the image corresponding to an arbitrarily placed object can be readily calculated. However, it must be remembered that, in order to determine the cardinal points in this way, not only must both the image and object be outside of the lens (i.e., in field-free space), but the same must be true of the focal points.

An alternative method of determining the imaging properties of a lens is to trace through the system, first, a ray having a small radial initial velocity leaving an object point on the axis which is followed until it intersects the axis, thus locating the image plane, and second, a ray from the object leaving with zero radial initial velocity from a point off the axis whose intersection with the image plane will give the magnification. With the aid of these two ray paths the four cardinal points may also be determined. If the image and object are not in the lens, the cardinal points determined by this second procedure can be used irrespective of whether or not they fall within the lens. If the image or object lies within the lens, the calculation applies to that particular object or image position only.

The differential ray equation rarely permits an analytical solution in practical cases. Again, as was found to be necessary in the solution of the Laplace equation, numerical approximations or graphical methods must be resorted to.

A very simple and rapid approximate method has been proposed by Richard Gans.* With a little care, this method is capable of an accuracy adequate for most practical cases.

The method consists of representing the potential along the axis by a series of straight-line segments, and applying the ray equation along the segments in turn.

* See Gans, reference 14.

Over any straight segment the second derivative $d^2\phi/dz^2$ is zero. The ray equation therefore becomes:

$$\frac{d^2r}{dz^2} = -\frac{1}{2\phi}\frac{dr}{dz}\frac{d\phi}{dz}.$$

This can be integrated twice to give:

$$\frac{dr}{dz}\sqrt{\phi} = C, \tag{4.31}$$

and

$$r = r_0 + \frac{2C(\sqrt{\phi} - \sqrt{\phi_0})}{\dfrac{d\phi}{dz}}, \tag{4.31a}$$

where r_0 and ϕ_0 are the radial position and potential at the beginning of the segment.

At the point where two segments meet, $d^2\phi/dz^2$ is infinite. Integrating Eq. 4.30 over the transition, the equation

$$\left(\frac{dr}{dz}\right)_2 - \left(\frac{dr}{dz}\right)_1 = -r\frac{\left(\dfrac{\partial\phi}{\partial z}\right)_2 - \left(\dfrac{\partial\phi}{\partial z}\right)_1}{4\phi} \tag{4.32}$$

is obtained. The subscript 1 indicates values before the break-point, and 2 those after the break-point.

Finally, where the segment is parallel to the axis, the solution becomes

$$r = r_0 + \left(\frac{dr}{dz}\right)_0 (z - z_0). \tag{4.33}$$

The method, because of its utility as a practical means for estimating the performance of any electron lens system for which the axial distribution is known, is worth illustrating by means of a simple numerical example.

The lens system to be considered consists of two equidiameter coaxial cylinders. The cylinders are assumed to have potentials equal to 2 and 12, respectively. Here the units of potential are quite arbitrary, the only thing of importance, as far as the lens properties are concerned, being the voltage ratio, which in this example is 6. The axial potential distribution of the system has been calculated by means of Eq. 4.35a given in the next section, and is tabulated in Table 4.1, in which the distance z is given in cylinder radii.

TABLE 4.1

z	ϕ	z	ϕ
−2.0	2.05	0.1	7.66
−1.8	2.10	0.3	8.88
−1.6	2.16	0.5	9.88
−1.4	2.27	0.7	10.62
−1.2	2.44	0.9	11.12
−1.0	2.70	1.1	11.45
−0.8	3.12	1.3	11.66
−0.6	3.72	1.5	11.79
−0.4	4.58	1.7	11.87
−0.2	5.71	1.9	11.93
−0.0	7.00		

It should be pointed out that the potential distribution for any desired voltage ratio can be obtained with the aid of the values given in Table 4.1 by adding an appropriate constant.

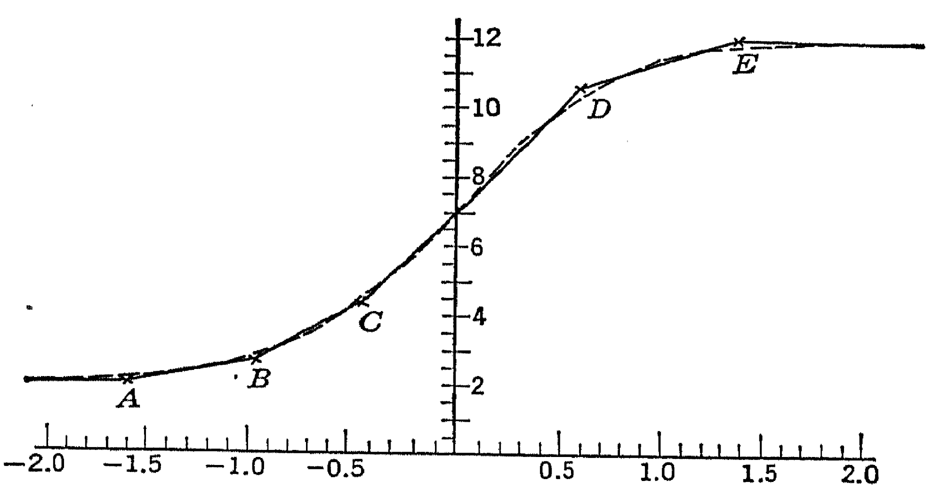


FIG. 4.12.—Axial Potential Distribution and its Line Segment Approximation for a Cylinder Lens.

The potential distribution is shown by the dotted curve of Fig. 4.12. In addition, this distribution is approximated in the figure by a series of six straight-line segments intersecting at points A, B, C, D, E. The intersections have the following values:

	z	ϕ
A.	−1.60	2.0
B.	−0.96	2.7
C.	−0.46	4.3
D.	0.60	10.6
E.	1.40	12.0

The slopes of the segments, and consequently the potential gradients, ϕ', represented by them, are:

Segment	ϕ'
∞A	0
AB	1.09
BC	3.20
CD	5.95
DE	1.75
$E\infty$	0

An electron leaves an object point on the axis at a distance -12 lens radii from the origin, with velocity corresponding to $\phi = 2$ and a slope of 0.096. It is required to find the conjugate image point, that is, the point at which the electron again intersects the axis.

The electron will move in a straight line until it reaches the first break-point at A. At this point, its radial distance is $r_A = 1.00$ and its slope $r_1{}'_A = 0.096$. Its slope at the other side of the break-point is given by Eq. 4.32:

$$r_2{}'_A = 0.096 - \frac{1.00 \times (1.09 - 0)}{4 \times 2.0}$$
$$= -0.040.$$

The value for C for the segment AB, given by Eq. 4.31, is:

$$C = -0.040 \times \sqrt{2.0} = -0.057.$$

Therefore, by Eq. 4.31:

$$r_1{}'_B = \frac{-0.057}{\sqrt{2.7}} = -0.034,$$

and from Eq. 4.31a:

$$r_B = 1.00 + \frac{-0.114(\sqrt{2.7} - \sqrt{2.0})}{1.09}$$
$$= 0.97.$$

The calculation continues in the same way for each succeeding break-point and segment. The results are given in Table 4.2.

TABLE 4.2

	r	$r_1{}'$	$r_2{}'$	C
A	1.00	0.096	−0.040	−0.057
B	0.97	−0.034	−0.022	−0.366
C	0.88	−0.18	−0.32	−0.657
D	0.62	−0.20	−0.14	−0.459
E	0.51	−0.13	−0.11	

From point E the electron continues in a straight line, with a slope to the axis of -0.11. Therefore, it intersects the axis at a distance $0.51/0.11 = 4.6$ radii from point E, or, in other words, the image point is 18.0 lens radii from the object point. A determination of the path by the more exact methods outlined in the next section gives the distance between the object and image point as 18.6 lens radii. Fig. 4.13 shows the exact path of the electron through the system, together with the

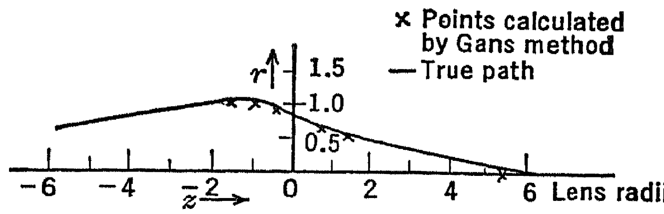


FIG. 4.13.—Electron Trajectory through Cylinder Lens.

points corresponding to the radial position at the break-points as calculated by the approximate method.

A closer approximation, of course, could have been obtained if a greater number of line segments had been used to represent the potential, particularly on the low-potential side of the lens.

4.9. Special Lens Systems. The electron-optical system consisting of two coaxial cylinders of equal diameter forms the basis of many practical lenses The properties of this configuration can, of course, be determined by mapping the axial potential distribution with the aid of a plotting tank and then tracing suitable rays by the method just discussed. However, when high accuracy is required, a mathematical solution may be advantageous.

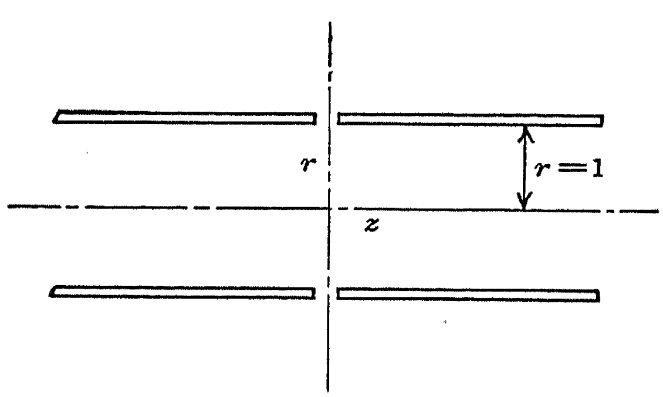


FIG. 4.14.—Coordinates for Electrostatic Cylinder Lens.

Fig. 4.14 illustrates, in cross-section, two semi-infinite coaxial cylinders, spaced a negligible distance apart. The axis of the cylinders will be taken as the z-axis of the cylindrical coordinate system, and the origin of the system will be located at the junction of the two cylinders.

The general solution of the Laplace equation for this type of configuration was discussed in the preceding chapter. By separating the variables, it was shown that the solution could be expressed in the form of the following integral:

$$\phi(r,z) = \int A(k)G(r,k)F(z,k)\,dk,$$

where:

$$F = ae^{ikz} + be^{-ikz},$$
$$G = cJ_0(ikr) + dN_0(ikr).$$

The requirement that the potential remain finite as z increases eliminates all terms with complex k. Furthermore, the condition that it be finite along the axis requires that the coefficients of the Neumann function be zero. Finally, since $\phi - (\phi_1 + \phi_2)/2$ is an odd function of z, the only trigonometric functions to be considered are sines.

Hence, the solution can be written as:

$$\phi(r,z) = \int_0^\infty B(k)J_0(ikr)\sin kz\,dk + \frac{\phi_1 + \phi_2}{2}. \tag{4.34}$$

The coefficient $B(k)$ can be found with the aid of boundary conditions for $r = 1$, the radius of the cylinders being taken as unit length; these are:

$$\phi(1,z) = \phi_1 \text{ for } z<0,$$
$$\phi(1,z) = \phi_2 \text{ for } z>0.$$

This evaluation leads to:

$$B(k)J_0(ik) = \frac{1}{\pi}\left\{\int_{-\infty}^0 -\frac{(\phi_2 - \phi_1)}{2}\sin kz\,dz + \int_0^\infty \frac{(\phi_2 - \phi_1)}{2}\sin kz\,dz\right\},$$

$$B(k) = \frac{(\phi_2 - \phi_1)}{\pi k J_0(ik)}(1 - \lim_{z\to\infty}\cos kz).$$

This value being placed in Eq. 4.34, the potential is found to be:

$$\phi(r,z) = \frac{1}{\pi}\int_0^\infty \frac{\phi_2 - \phi_1}{kJ_0(ik)}J_0(ikr)\sin(kz)\,dk + \frac{\phi_1 + \phi_2}{2}.$$

The axial potential and its first two derivatives are:

$$\phi(0,z) = \frac{1}{\pi}\int_0^\infty (\phi_2 - \phi_1)\frac{\sin kz}{kJ_0(ik)}dk + \frac{\phi_1 + \phi_2}{2}, \tag{4.35a}$$

$$\phi'(0,z) = \frac{1}{\pi}\int_0^\infty (\phi_2 - \phi_1)\frac{\cos kz}{J_0(ik)}dk, \tag{4.35b}$$

$$\phi''(0,z) = -\frac{1}{\pi}\int_0^\infty (\phi_2 - \phi_1)\frac{k\sin kz}{J_0(ik)}dk. \tag{4.35c}$$

These integrals cannot be solved analytically, but must be evaluated

by quadrature, which, though laborious, is a fairly straightforward procedure. Numerical values of potential thus obtained are used to calculate the electron paths from the ray equation.

The solution of the ray equation is expedited by the substitution:

$$c = -\frac{1}{r}\frac{dr}{dz}.$$

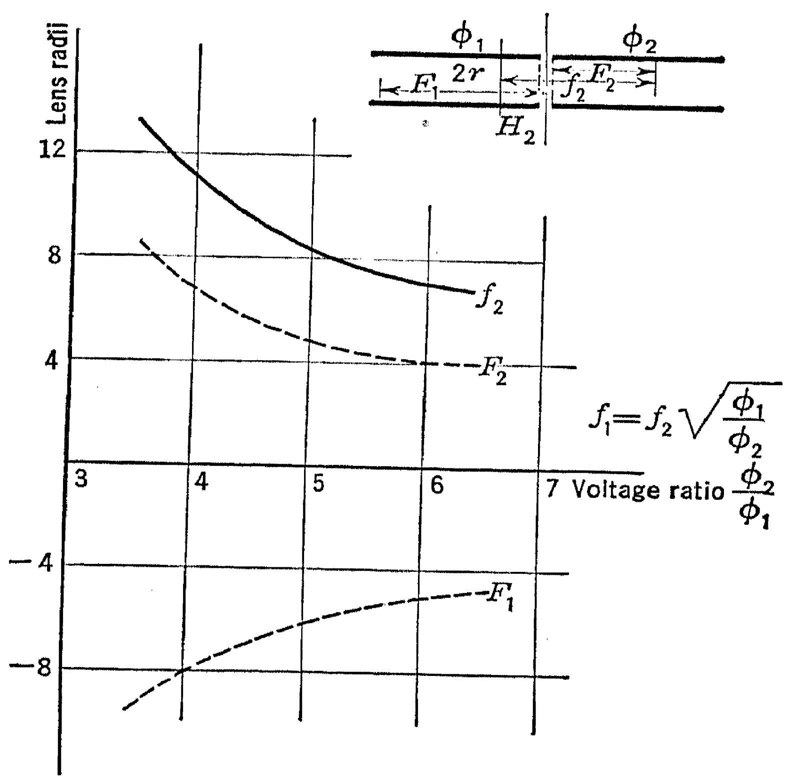


FIG. 4.15.—Focal Length and Position of Focal Points of Cylinder Lens as Function of Voltage Ratio (After Epstein).

The function c is the convergence of the ray. It is thus named because it is the reciprocal of the distance along the axis from the point at which c is determined to the intersection of a rectilinear extension of the ray with the axis.

This substitution being made, the ray equation (4.30) becomes:

$$\frac{dc}{dz} = c^2 - \frac{c\phi'}{2\phi} + \frac{\phi''}{4\phi}. \tag{4.36}$$

The evaluation of the equation must be continued until the ray is substantially a straight line, or until it passes through the image plane.

The curves reproduced in Fig. 4.15 locate completely the four cardinal points for various values ϕ_1/ϕ_2 in the type of system just described. All distances are given in terms of lens radii.

Besides the coaxial cylinder lens using cylinders of the same diameters,

an infinite number of other lenses can, of course, be formed by using cylinders of unequal diameters. The effect of this difference is to alter to some extent the position of the cardinal points. From the standpoint of Gaussian dioptrics, there is rarely any advantage in choosing cylinders of different diameters. Empirical data, however, indicate that certain of the image defects can be considerably reduced by a proper selection of diameter ratios.

These electron-optical systems will be referred to again in the discussion of the electron gun.

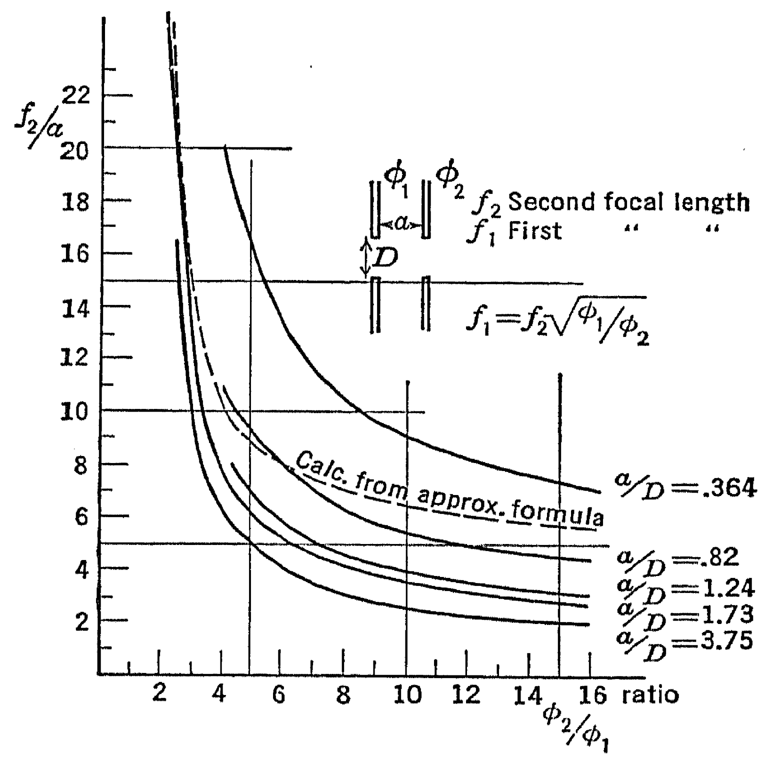


FIG. 4.16.—Focal Length of Double-Aperture Lenses (After Polotovsky).

4.10. Aperture Lenses. Another very important class of lenses are those formed by a pair of apertured conductors at different potentials. This lens is used extensively both in the electron gun and in electron imaging devices such as the microscope.

The simplest lens of this type has already been discussed. The results obtained for this restricted case, however, have very limited application. They can be used only where the potential difference of the apertures is small so that the focal lengths are large compared to the spacing between the apertures and to their diameters.

A determination of the first-order-image properties of aperture lenses has been made by Polotovsky by electrolytic and graphical methods. Fig. 4.16 reproduces some of his results. These curves give the second focal length for various voltages, and separation to diameter ratios. For comparison, Eq. 4.28 for the simplified case is included in this figure.

4.11. Cathode Lens Systems. In many practical electron-optical lenses, the electrons enter the system with essentially zero velocity. Such a lens is to be found in the electrostatically focused image tube (which, in connection with the Iconoscope, is discussed in Chapter 11). This type of lens, shown in Fig. 4.17, differs quite radically in its properties from the lenses discussed so far.

The potential at the point where the electron enters the system is zero. In consequence, the electron ray originates in a region of zero index of refraction. Because of this, the first focal length is zero and the first principal plane is located at the object.

Under these conditions, if the electrons were actually emitted from the object, that is, the cathode, with zero velocity, the image plane would be completely indeterminate. However, the assumption of a small radial initial velocity suffices to determine the position and, with it, the

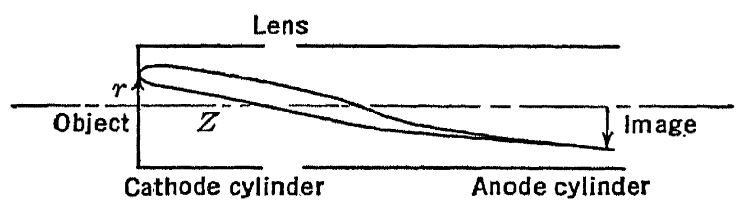


FIG. 4.17.—Cathode Lens.

magnification of the image. In practice, both the radial and axial components of the initial velocity are far from negligible, and are responsible for certain aberrations discussed in section 4.14.

One of the simplest lenses of this type consists of two coaxial cylinders of equal diameter, one infinite in extent, the other terminated by a flat plane. The terminal plane is assumed to be the electron object. In a conventional image tube, for example, this plane would be a photoelectric cathode. The cathode surface is conducting and is electrically connected to the cathode cylinder, these elements being at zero potential. For convenience, the lens will be said to be located at the junction of the two cylinders, although actually it extends from the cathode to a distance of several cylinder radii beyond the junction. The geometry of the lens can be seen from Fig. 4.17.

The calculation of the potential distribution proceeds exactly as in the previous example of the two-cylinder lens. It is expedient here to use cylindrical coordinates with their origin at the cathode. The general solution of the Laplace equation, which can be evaluated from the boundary conditions of the lens, leads to the following integrals, giving the potential along the axis together with the first and second derivatives:

$$\phi(0,z) = \frac{2}{\pi}\int_0^\infty \phi_1 \frac{\cos ku \sin kz}{kJ_0(ik)}dk, \tag{4.37a}$$

$$\phi'(0,z) = \frac{2}{\pi}\int_0^\infty \phi_1 \frac{\cos ku \cos kz}{J_0(ik)}dk, \tag{4.37b}$$

$$\phi''(0,z) = -\frac{2}{\pi}\int_0^\infty \phi_1 \frac{k \cos ku \sin kz}{J_0(ik)}dk. \tag{4.37c}$$

As before, these integrals must be evaluated by quadrature.

The numerical values of the potential thus obtained can, with the aid of the ray equation, be used to obtain the lens properties. In this instance, the convergence function, c, becomes infinite at $z = 0$ and thus cannot be used. However, the substitution of the function:

$$b = \frac{1}{2z} - \frac{1}{r}\frac{dr}{dz},$$

leading to:

$$\frac{db}{dz} = b^2 - b\left(\frac{1}{z} + \frac{\phi'}{2\phi}\right) + \frac{\phi''}{4\phi} + \frac{1}{2z}\left(\frac{\phi'}{2\phi} - \frac{1}{2z}\right),$$

will facilitate the solution by reducing the ray equation to a first-order differential equation.

Numerical determinations of two electron paths will locate the image and give its magnification. In Fig. 4.18 the potential and its first two derivatives are shown, together with two electron trajectories. It should be noted that the latter are independent of the applied potential, ϕ_1, and are a function of the cathode-to-lens distance, u, only. The relations between image distance, object distance, and magnification are given in Fig. 11.9 on page 317.

It is often desirable to arrange the lens in such a way that the focal length and image position can be varied electrically. This can be done by making the cathode cylinder of resistive material so that an axial potential gradient can be established on it between cathode and lens. In practice, it is convenient to divide the cathode cylinder into a number of rings having equal potential steps between them. Experimentally, this is found to simulate the resistive cathode cylinder qualitatively and quantitatively.

The performance of this system can be calculated by the methods already given. The results of such a determination are of value in many design problems.

The magnification of the system just described is a function of the image and object distance alone and cannot be changed in a given tube configuration. It is possible, by incorporating a third element, to make

a lens system which has a variable magnification. The additional electrode may be in the form of an aperture or a short cylinder placed between the anode and cathode cylinders. An electron-optical system of this type is illustrated in Fig. 4.19. For a given tube structure and overall

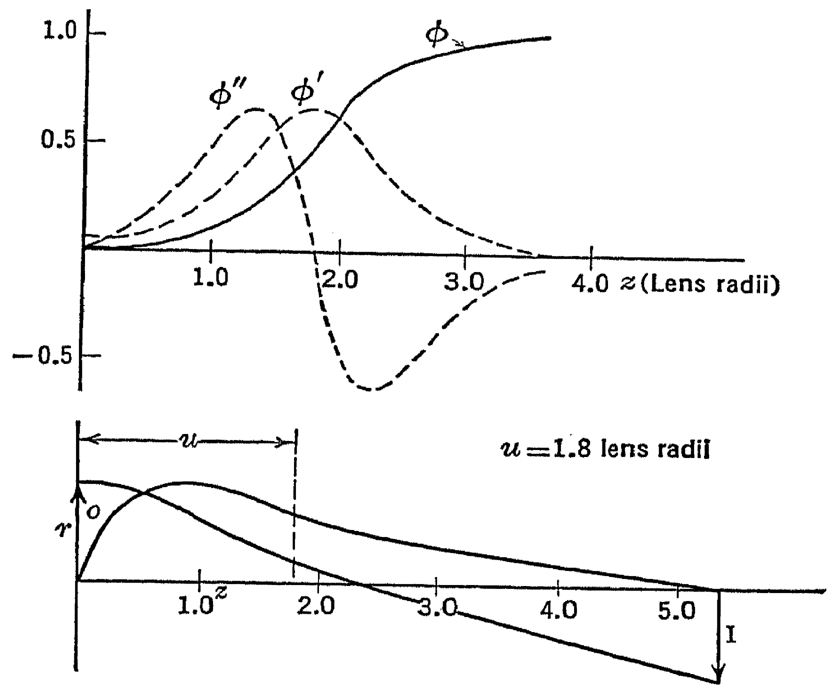


FIG. 4.18.—Axial Potential Distribution and Electron Trajectories in a Cathode Lens.

voltage, the magnification is varied by changing the potential ϕ_3 on the extra electrode and refocusing the image by the potential ϕ_2 across the cathode cylinder.

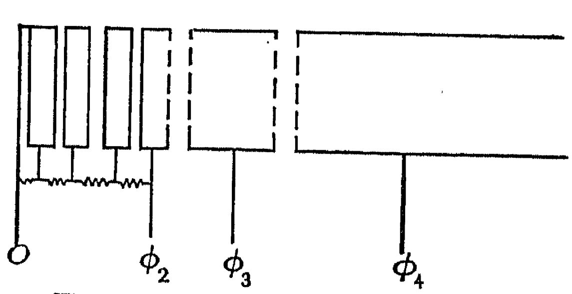


FIG. 4.19.—An Electron Lens System Arranged to Give Variable Magnification.

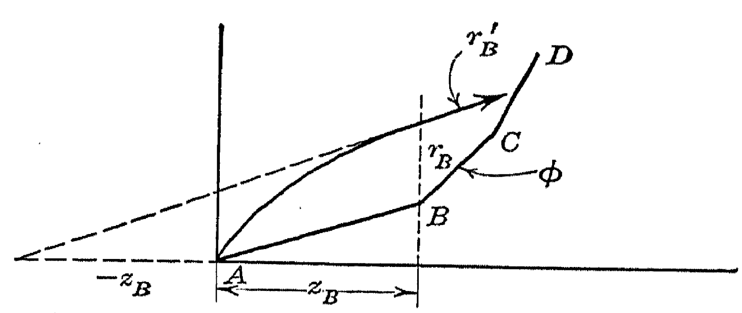


FIG. 4.20.—Determination of Electron Trajectory Near Cathode in Gans Method of Ray Tracing.

Before leaving the subject of cathode lens systems, some mention should be made of the procedure required when they are to be investigated by the approximate method described in section 4.8. In order to determine the image position, it is necessary to assume an initial radial velocity. Referring to Fig. 4.20, the axial potential distribution of a cathode lens system is represented by a series of straight-line segments.

The electron is emitted from the object point, A, with a small radial initial velocity, into a region of uniform field as represented by the first segment, AB. The trajectory of the electron is a parabola. If at the end of the first segment its radial distance from the axis is r_B, then, from the properties of the parabola, the slope of its trajectory will be $r_B/2z_B$, where z_B is its axial position. Therefore, the electron will enter the first break-point on a path described as follows:

$$\left.\begin{aligned} \left(\frac{dr}{dz}\right)_1 &= \frac{r_B}{2z_B}, \\ r &= r_B, \\ \phi &= \phi_B. \end{aligned}\right\} \tag{4.38}$$

From this point on, the procedure is the same as for any lens system. The choice of the radial separation r_B is arbitrary, and does not affect the determination of the image position.

4.12. The Magnetic Lens. Thus far the discussion has been limited to electrostatic lens systems. Magnetic lenses, however, rank as at least

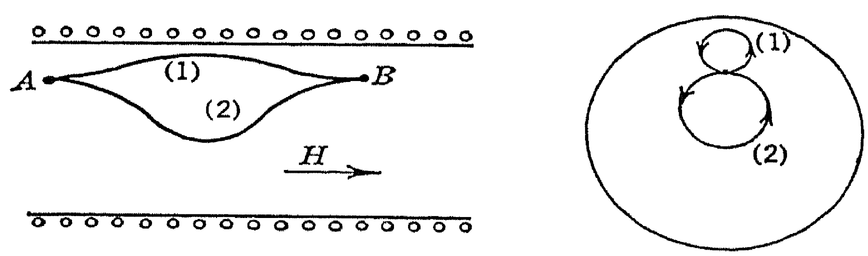


Fig. 4.21.—Electron Paths in a Uniform Magnetic Field.

equal in importance. From a theoretical standpoint they are very much more complicated, because the principal electron paths are not confined to a plane.

The simplest magnetic lens consists of a uniform magnetic field parallel to the axis of the system and pervading all the region between anode and cathode.

An electron which leaves the emitter parallel to the axis is subject to no force from the magnetic field. The time required by this electron to traverse the distance between anode and cathode is given by the integral:

$$t_1 = \int_0^L \frac{dz}{v_z}, \tag{4.39}$$

where L is the total distance and v_z the axial velocity. When, however, the particle leaves with a radial initial velocity, v_r, it will experience a force

$$F = Hev_r,$$

where H is the magnetic field strength. Since the direction of this force is at right angles to the transverse velocity and to the field, the electron path will be curved. The projection of this path on a plane normal to the axis is a circle, the radius ρ being given by the relation between the inward acceleration and the centripetal force, i.e.:

$$\frac{mv_r^2}{\rho} = Hev_r,$$

$$\frac{v_r}{\rho} = \frac{e}{m}H.$$

The circumference of this circle will be $2\pi\rho$, and, therefore, the time t_2 required to traverse the circle will be:

$$t_2 = \frac{2\pi\rho}{v_r}$$

$$= \frac{2\pi m}{eH}. \tag{4.40}$$

This time is independent of the initial radial velocity. If t_2 in Eq. 4.40, or any multiple thereof, is equal to t_1 in Eq. 4.39, all electrons leaving a point on the source will come together at a point in the image plane, regardless of their initial radial velocity. It should be noticed that the image, unlike that formed by an electrostatic lens, is erect.

The magnetic lens representing the other extreme is also amenable to simple calculation. This lens is the limiting case of the short lens. Here the magnetic field strength is negligible except over a distance which is small compared to the distance between object and image. The region over which the field is appreciable will be termed the lens. Furthermore, throughout this region it will be assumed that both the potential, ϕ_0, and the radial distance, r_0, of the ray from the axis are constant. The slope of the ray, dr/dz, of course, changes. On the basis of these assumptions it can be shown that:

$$\left(\frac{dr}{dz}\right)_A - \left(\frac{dr}{dz}\right)_B = \frac{er_0}{8m\phi_0}\int_A^B H^2(z)dz, \tag{4.41}$$

where $H(z)$ is the axial magnetic field, and the slopes with subscripts A and B are those of the ray as it enters and leaves the lens region. If s and s' are the object and image distance, respectively, for an object point on the axis, it follows that:

$$\left(\frac{dr}{dz}\right)_A - \left(\frac{dr}{dz}\right)_B = \frac{r_0}{s} + \frac{r_0}{s'}.$$

Therefore, it follows that:

$$\frac{1}{f} = \frac{e}{8m\phi_0}\int_A^B H^2(z)dz. \tag{4.42}$$

Its magnification is:

$$m = -\frac{s'}{s}.$$

This type of system differs from that consisting of a uniform magnetic field in that the image is inverted.

In general, the practical magnetic lens lies between the two extremes just described. It is capable of forming a real image, and the image will be rotated through an angle whose magnitude depends upon the configuration and magnitude of the field.

Normally in magnetic lenses the variation in electrostatic potential throughout the system is such that its converging action is negligible compared to that due to the magnetic field, and the differential ray equation for a paraxial ray is:

$$\frac{d^2r}{dz^2} = -\frac{eH^2}{8m\phi}r. \tag{4.43}$$

The image is rotated through an angle θ, where

$$\theta = \sqrt{\frac{e}{8m\phi}}\int H dz. \tag{4.43a}$$

An exact solution of Eq. 4.43 is difficult and, usually, impossible.

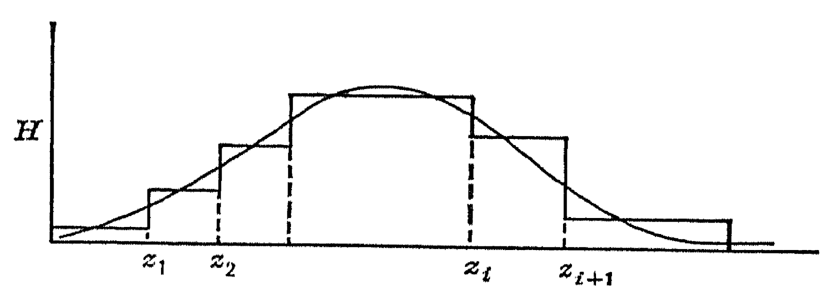


Fig. 4.22.—Approximate Representation of Axial Magnetic Field.

An approximate method of solution similar to that used for the electrostatic lens has been proposed by E. G. Ramberg.

To illustrate this method it is assumed that the axial distribution of a field has been measured or estimated, leading to the smooth curve in Fig. 4.22. The actual distribution is approximated by a series of step segments as shown in the figure.

The solution of Eq. 4.43 over any segment, for example, that between z_i and z_{i+1}, is:

$$r = r_i \cos \sqrt{\frac{eH^2}{8m\phi}}(z - z_i) + \left(\frac{dr}{dz}\right)_i \frac{\sin \sqrt{\frac{eH^2}{8m\phi}}(z - z_i)}{\sqrt{\frac{eH^2}{8m\phi}}}, \qquad (4.44)$$

where ϕ is the potential at the segment, H the magnetic field, r the radial distance to the ray, and $z_i < z < z_{i+1}$. Both r and dr/dz are continuous at the break-points, so that the solutions for the individual segment given by Eq. 4.44 join smoothly.

With the aid of the initial conditions, the substituted field distribution, and Eq. 4.44, the path of the ray (or, more exactly, the radial displacement) can be traced through the lens, just as is done for the electrostatic system. Two ray paths are sufficient to determine the first-order-image properties of the magnetic lens.

Finally, considering the most general case where the lens action is due both to magnetic and electrostatic fields, the differential ray equation becomes:

$$\frac{d^2r}{dz^2} = -\frac{1}{2\phi}\frac{dr}{dz}\frac{d\phi}{dz} - r\left(\frac{1}{4\phi}\frac{d^2\phi}{dz^2} + \frac{eH^2}{8m\phi}\right), \qquad (4.45)$$

where the symbols have the previously assigned meanings. The rotation of the image is given by

$$\theta = \sqrt{\frac{e}{8m}} \int \frac{H}{\sqrt{\phi}} dz.$$

4.13. Image Defects. The theory thus far developed has dealt with images formed by rays very close to the axis of the lens. In other words, it deals with rays making angles with the axis which are sufficiently small that only the first term of the sine expansion need be used. However, because the first-order theory deals with such a small portion of the lens and object, it tells nothing about the quality of the image.

The theory has been extended to take into account the next term of the expansion. This third-order theory, developed by Ludwig von Seidel and bearing his name, leads to a series of five correction terms which vanish when the oblique rays are deflected in the same manner as the paraxial rays. When one or more of these terms differ from zero, the oblique rays from the object do not converge in the image points indicated by the laws governing the paraxial rays and the image is unsharp or deformed; the system is then said to have aberrations. Corresponding

to the five terms in the third-order theory, there are five aberrations from which a system may suffer. These are:

Spherical aberration.
Astigmatism.
Coma.
Curvature of the image field.
Distortion of the image.

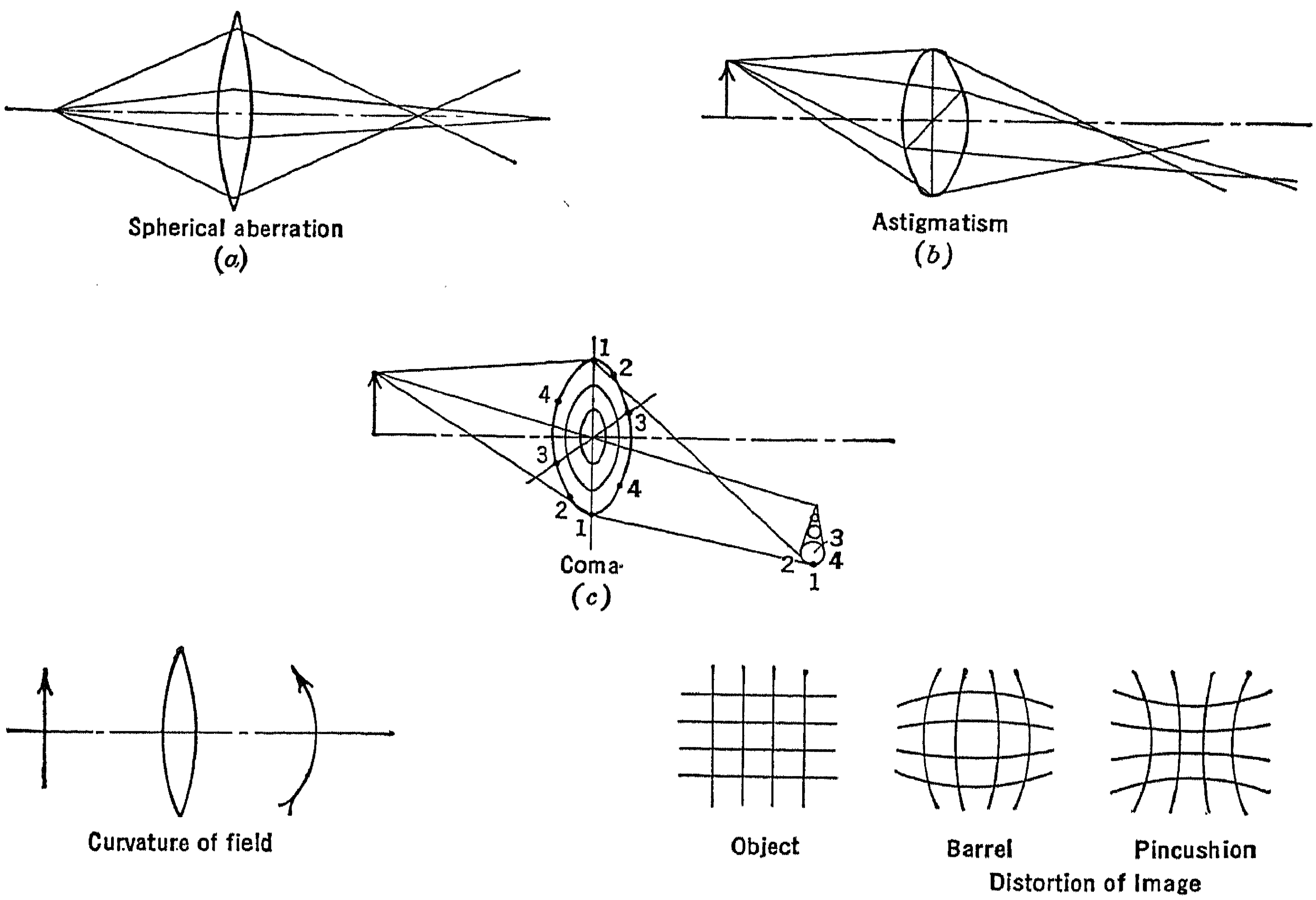


FIG. 4.23.—Nature of the Five Third-Order Aberrations.

In addition to these five third-order aberrations there is a sixth due to the variations in the initial velocities of the electrons as they leave the object. This is known as chromatic aberration because of its similarity to optical chromatic aberrations caused by the variation of the index of refraction with the wavelength of light.

These aberrations play an important role in electron imaging and, hence, merit consideration.

Spherical aberration occurs when rays leaving the object at the axis and passing through the outer portions do not converge at the paraxial image point. This is illustrated in Fig. 4.23*a*.

Astigmatism. The images of object points lying at a distance from the

axis of the system suffer from an additional defect resulting from the fact that rays in a plane containing the axis converge on a different point from those in a plane perpendicular to this plane. This defect, known as astigmatism, is shown in Fig. 4.23*b*.

Coma. Even in the absence of the two afore-mentioned aberrations the image of a point off the axis may not be sharp, but will be a comet-shaped area whose vertex coincides with the first-order image point. Because of the form of the image this defect is known as coma.

Curvature of the Image Field. Again leaving chromatic defects out of consideration, every point of the object will be sharply imaged after correction has been made for spherical aberration, coma, and astigmatism. However, the image points may not lie in a plane unless correction for the curvature of the image field has been made.

Distortion of the Image. After the four corrections mentioned above have been made, a fifth aberration may be present, consisting of a nonuniformity of the magnification or a twist of the image.

4.14. Aberrations. The above classification may be derived from a consideration of the symmetry conditions of image formation; hence, it is

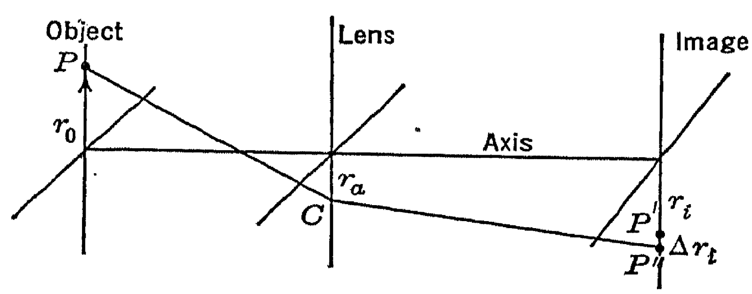


FIG. 4.24.—Object, Lens, and Image Planes of an Electron-Optical System.

applicable to optical and electron lens systems. The exact derivation of these aberrations is beyond the scope of this chapter, but the dependence of the aberrations on the various elements of the system can be shown in the following simplified way.

Fig. 4.24 represents a plane through the lens system which includes the axis. The lens itself is confined to a narrow region as indicated on the figure. An object point, P, is located at a distance r_o from the axis and the corresponding image point, P', at r_i. In order that the image be faithful, r_i must be proportional to r_o. This condition is fulfilled by the first-order image. Next, consider the ray PCP'', which originates at the object point at r_o and meets the image plane at r_i'. It is evident that r_i' is a function of r_o and r_a, and that by expansion this dependence can be expressed as a power series with constant coefficients. This expression of dependence will be symbolized by:

$$r_i' \rightarrow r_o,\ r_a,\ r_o^2,\ r_a^2,\ r_or_a,\ r_o^3,\ r_o^2r_a,\ r_or_a^2,\ r_a^3.$$

The first-order image, by definition, is such that, when r_o and r_a are small, the ray PCP'' will meet the image point. Therefore, the aberration which is given by:

$$\Delta r_i = r_i' - r_i$$

is not dependent upon the first power of r_o or r_a. Furthermore, the cylindrical symmetry of the system makes it impossible for r_i to depend on any term of even order. Thus there remains the aberration of the third order whose dependence may be expressed as follows:

$$\Delta r_i \rightarrow r_o^3,\ r_o^2 r_a,\ r_o r_a^2,\ r_a^3.$$

The five aberrations are classified according to the terms of this expansion appearing in them. As the coefficients of these terms become zero, the corresponding aberration vanishes.

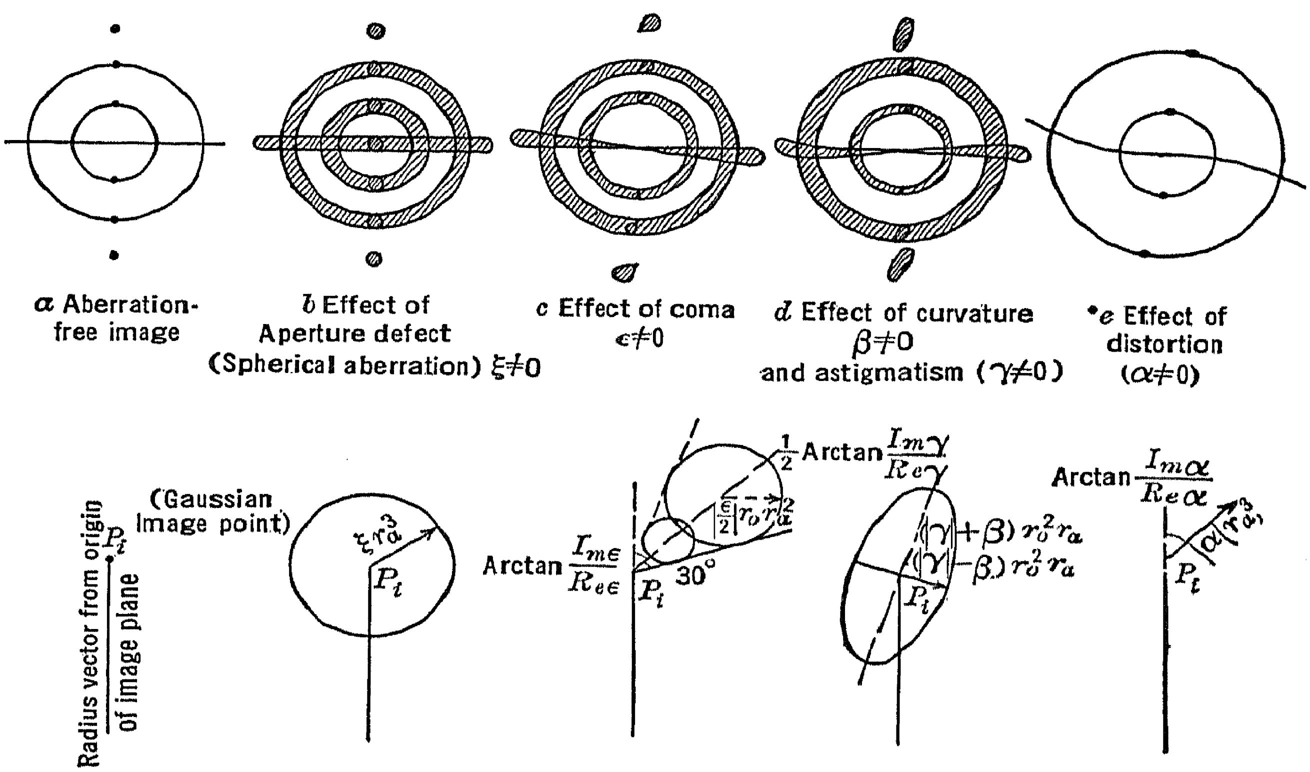


FIG. 4.25.—The Effects of the Individual Geometric Aberrations on an Electron Image.

Spherical aberration is that corresponding to r_a^3; it is therefore proportional to the cube of the diameter of the ray bundle of electrons as they go through the lens region. Since r_o does not appear in this aberration, it exists for object points on the axis to the same extent as for those off the axis.

The term with $r_o r_a^2$ indicates the extent of coma present. Since r_o appears in this term, coma vanishes for a point on the axis. Both curvature

of the image field and astigmatism are the result of non-negligible coefficients of terms in $r_o^2 r_a$. Like coma, these are field aberrations.

The final term in r_o^3 is that giving the distortion of the image.

It will be clear from a consideration of such aberrations as coma and astigmatism that a representation in a plane containing the axis is not sufficient to describe them, since both require a consideration of rays which do not lie in this plane. A coordinate corresponding to a transverse displacement must be introduced in a complete analysis. However, this approximate survey will serve to indicate the nature of the third-order image defects, which are shown in detail in Fig. 4.25.

The procedure for calculating the image defects of an actual system is extremely difficult and laborious. The usual method is to determine the first-order image as has been described, and then to calculate the separation between the intersection of a non-paraxial ray with the image plane and the corresponding image point.

For more complete discussions of the analytical methods of dealing with these aberrations, the reader is referred to the published work of Scherzer and Glaser.*

From a practical standpoint, information as to the magnitude and method of correction of these aberrations is very important, because one of the major purposes of electron optics is the design of systems capable of producing good images. Unfortunately, only a very limited amount of data on this subject is available. Some general conclusions can be drawn which are helpful in the laying out of practical electron lens systems.

It can be shown that it is impossible to eliminate spherical aberration completely in electrostatic, magnetic, or combined systems. In other words, an aplanatic lens does not exist in electron optics. Furthermore, the spherical aberration for a combined electric and magnetic lens cannot be made appreciably less than for either type alone. Immersion lenses having a photoelectric cathode as object, such as are used in the image tube, in general require the focusing of extremely narrow ray bundles, and hence this aberration will be small. Spherical aberration, however, is a very important defect in the electron gun and is one of the limiting factors in producing a small spot. In this connection this aberration has been studied for concentric cylindrical lenses by Epstein and by Gundert.†

Curvature of field and astigmatism, as was shown, have the same dependence upon object position and radius of the lens aperture. No practical lens having a plane object has been found by either theoretical or experimental methods which is free from these defects, except those consisting of a uniform electric and magnetic field. Furthermore, the image

* See references 2, 9, and 11.

† See Epstein, reference 16, and Gundert, reference 20.

plane is invariably concave toward the lens, the tangential image having the greater curvature. These defects are dependent on the first power of the aperture and the square of the distance of the object point from the axis, and hence are very important in the image tube, despite the small aperture of the electron beams, as the object field is large. The formation of the sagittal and tangential images in a flat cathode image tube is shown in Fig. 4.26. It has been found possible to correct these two defects in the image tube to a large extent by curving the cathode. Fig. 11.10, page 318, illustrates the image formed by an uncorrected and corrected system, together with the electron-optical systems producing them. Since these aberrations are field aberrations which vanish for points on the axis, they do not enter into the design of the electron gun.

Coma can be made to vanish in a system where the electric field is symmetrical about a plane midway between the object and image, and the magnetic field anti-symmetric. In the image tube, where the electron

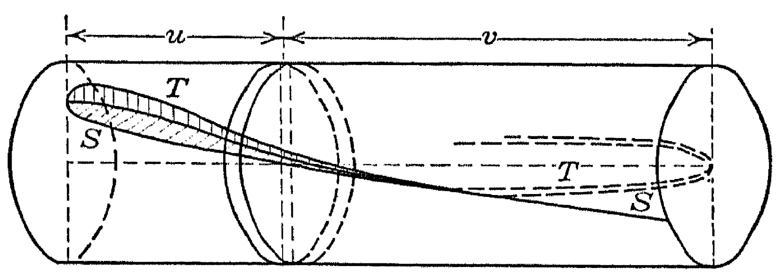


Fig. 4.26.—Curvature of Field and Astigmatism in a Flat-Cathode Electrostatic Image Tube.

has zero velocity at the cathode, this type of correction obviously is impossible as the electron would have to have zero velocity at the image surface as well. This aberration vanishes on the axis and is, consequently, of no importance in the design of an electron gun. Experience indicates that this defect is of small importance in the photoelectric image tube.

Distortion, the last of the five Seidel aberrations to be considered, can be divided into two types: radial distortion and rotational distortion. The former alone is present in the electrostatic lens. When it has the form of increasing magnification with radial image distance, it leads to pincushion distortion, the type most prevalent in electron lenses. The converse variation of magnification leads to barrel distortion. Both rotational and radial distortion may be present in magnetic systems. Rotational distortion occurs when the rotation of the image point about the axis of the system varies with its radial position.

A consideration of the effect of variations in the initial velocities on the image shows that achromatic electron lenses are impossible; however, the use of an electron mirror permits complete correction for two velocities.

The magnitude of this aberration for electron lenses will depend upon the ratio of initial velocity (in volts) to the overall voltage, and upon the configuration of the lens. For simple image tube types, the diameter of the circle of aberration can easily be estimated. The three configurations illustrated in Fig. 4.27 have the following values for Δ, the diameter of the circle of aberration:

(*a*) simple accelerating field: $\Delta = 4L(V/\phi)^{1/2}$,
(*b*) uniform magnetic and electric fields superposed: $\Delta = 2L(V/\phi)$,
(*c*) image tube with electric or magnetic lens: $\Delta = 2m(V/F)$,

which, for a short magnetic lens when the whole field is applied between the lens and object, becomes:

$$\Delta = \frac{2m}{2m+1} L \frac{V}{\phi}.$$

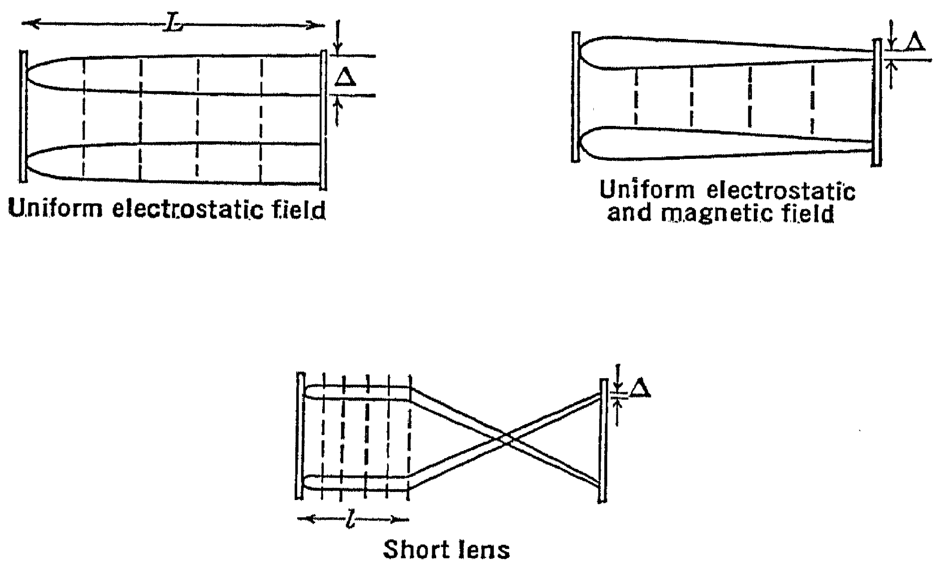


Fig. 4.27.—Chromatic Aberration for Three Simple Image Tubes.

In the above equations the symbols have the following significance: L is the length of the tube, m the magnification of the image, V the initial velocity in electron volts, F the field strength at the cathode, and ϕ the overall applied voltage.

The dependence of this aberration on the field strength at the cathode is apparent from the above relations. It is evident that high field strengths are required to reduce chromatic image defects.

The preceding treatment of electron optics and, in particular, the discussion of aberrations have necessarily been very much simplified and abbreviated since, although important in the development of a television system, electron optics represents a very small part of the whole field.

REFERENCES

1. E. BRÜCHE and O. SCHERZER, "Elektronenoptik," Julius Springer, Berlin, 1934.
2. H. BUSCH and E. BRÜCHE, "Beiträge zur Elektronenoptik," J. A. Barth, Leipzig, 1937.
3. I. G. MALOFF and D. W. EPSTEIN, "Electron Optics in Television," McGraw-Hill, New York, 1938.
4. H. BUSCH, "Cathode Ray Paths in Axially Symmetric Fields," *Ann. Physik*, Vol. 81, pp. 974–993, 1926.
5. M. KNOLL and E. RUSKA, "Geometric Electron Optics," *Ann. Physik*, Vol. 12, pp. 607–640, 641–661, 1932.
6. J. PICHT, "Theory of Geometric Optics for Electrons," *Ann. Physik*, Vol. 15, pp. 926–964, 1932.
7. E. BRÜCHE, "Fundamentals of Geometrical Electron Optics," *Z. tech. Physik*, Vol. 14, pp. 49–58, 1933. See, also, *Z. Physik*, Vol. 78, pp. 26–42, 1932.
8. O. SCHERZER, "Theory of Electric Electron Condensing Lenses," *Z. Physik*, Vol. 80, pp. 193–202, 1933.
9. W. GLASER, "Geometrical Optics of Electron Rays," *Z. Physik*, Vol. 80, pp. 451–464, 1933.
10. V. K. ZWORYKIN, "Electron Optics," *J. Franklin Inst.*, Vol. 215, pp. 535–555, 1933.
11. O. SCHERZER, "Errors of Electron Lenses," *Z. Physik*, Vol. 101, pp. 593–603, 1936.
12. W. ROGOWSKI, "Aberrations of Electron Images," *Arch. Elektrotech.*, Vol. 31, pp. 555–593, 1937.
13. L. H. BEDFORD, "Electron Lens Formulas," *Phys. Soc. Proc.*, Vol. 46, pp. 882–888, 1934. See, also, C. J. DAVISSON and C. J. CALBICK, *Phys. Rev.*, Vol. 38, p. 585, 1931; and Vol. 42, p. 580, 1932.
14. R. GANS, "Electron Paths in Electron Optics," *Z. tech. Physik*, Vol. 18, pp. 41–48, 1937.
15. L. S. POLOTOVSKI, "Studies on Electron Optical Systems," *Izveztia Electroprom. Slabovo Toka*, No. 5, pp. 22–35, and No. 6, pp. 13–27, 1934.
16. D. W. EPSTEIN, "Electron Optical Systems of Two Cylinders," *Proc. I. R. E.*, Vol. 24, pp. 1095–1139, 1936.
17. G. A. MORTON and E. G. RAMBERG, "Electron Optics of an Image Tube," *Physics*, Vol. 7, pp. 451–459, 1936.
18. E. G. RAMBERG, "Simplified Derivation of General Properties of an Electron Optical Image," *J. Optical Soc., Am.*, Vol. 29, pp. 79–83, 1939.
19. W. HENNEBERG and A. RECKNAGEL, "Chromatic Errors of Electron-Optical Systems," *Z. tech. Physik*, Vol. 16, pp. 230–235, 1935.
20. E. GUNDERT, "Demonstration of the Aberrations of Electron Lenses," *Physik. Z.*, Vol. 38, pp. 462–467, 1937.
21. B. VON BORRIES and E. RUSKA, "Electron Super Microscope," *Wiss. Veröffentl. Siemenswerke*, Vol. 17, No. 1, 1938.

CHAPTER 5

VACUUM PRACTICE

Nearly all work in electronics involves the use of one or more evacuated envelopes of either glass or metal. Therefore, it will not be amiss to devote at least a little space to a discussion of vacuum technique.

In practice, the degree of vacuum in a vessel is quantitatively expressed in terms of the pressure of the residual gas. Two frequently used units are the barye and the millimeter of mercury. The pressure expressed in baryes, or bars, is numerically equal to the number of dynes exerted by a medium on an area of 1 square centimeter. (Note: In meteorological work the bar is sometimes defined as 10^6 dynes per cm^2, or a pressure of approximately 1 atmosphere.) Expressed in millimeters of mercury, it refers to the height of the mercury column under the influence of gravity which can be supported by the gas pressure. One atmosphere corresponds to 760 mm Hg.

For most work in electronics the vacuum required is of a fairly high order, but by no means close to the best obtainable. The best vacuum that can be obtained, if every known precaution is taken, is about 10^{-9} mm Hg. For extremely accurate electronic measurements, such as are involved in the measurement of the work functions of pure metals, the pressure must be of the order of 10^{-7} to 10^{-8} mm Hg. For the electronic devices described in this book, pressures in the neighborhood of 10^{-6} mm are adequate, and often pressure above 10^{-5} mm can be tolerated in so-called hard tubes. Certain tubes involving a gas discharge require pressures of 10^{-2} mm to 10 mm. These figures are not to be taken as exact, but are merely given for the purpose of orientation.

At the pressures used in electronic devices, that is, 10^{-6} mm Hg, the space in the tube is far from empty of molecules. Actually, there are between 3 and 4×10^{10} molecules in each cubic centimeter. In spite of the great number of molecules present the average distance a molecule must travel before it strikes another is very large. This "mean free path" depends upon the kind of gas in the tube; for example, in air at this pressure it is about 6.5×10^3 cm, for hydrogen 12×10^3 cm, and for helium 19×10^3 cm. At constant temperature the "mean free path" is approximately inversely proportional to the pressure of a gas.

In addition to the vessel under exhaust, the conventional vacuum

system consists of a high-vacuum pump, such as a diffusion pump or a molecular pump, working into a fore-pump which reduces the pressure from atmospheric to 10^{-2} or 10^{-3} mm Hg. Provision is usually made for measuring the pressure on the high vacuum side in the form of a suitable gauge. Often the system is provided with additional elements, such as a stopcock and reservoir between the fore-pump and high-vacuum pump, a freezing trap on the high-vacuum side of the system to remove condensable vapors, and the auxiliary equipment needed in processing the

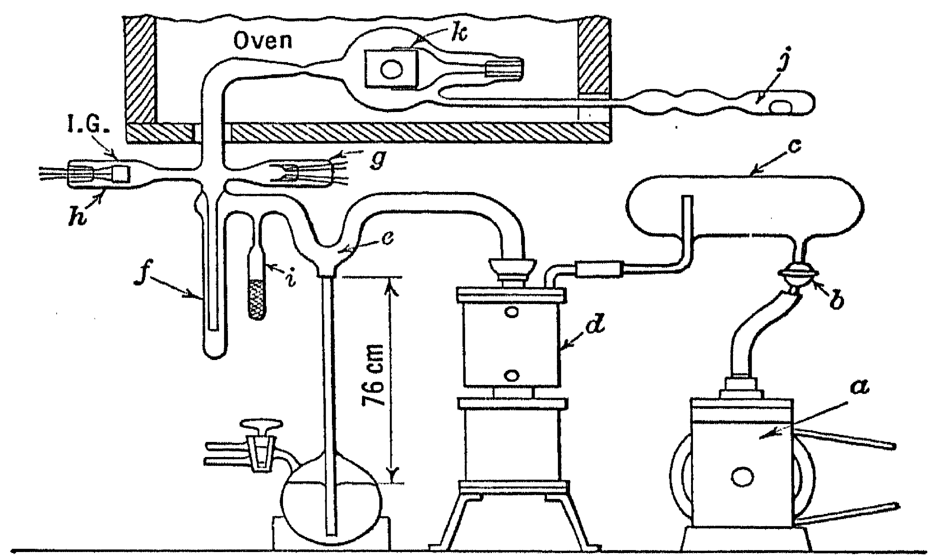


FIG. 5.1.—Schematic Diagram of a Typical Vacuum System.

tube being pumped. Fig. 5.1 shows a typical system schematically. The system illustrated is composed of the following elements:

(*a*) Fore-pump.
(*b*) Stopcock.
(*c*) Fore-vacuum reservoir.
(*d*) Mercury diffusion pump.
(*e*) Mercury cutoff.
(*f*) Liquid-air trap.
(*g*) Thermocouple gauge (high-pressure).
(*h*) Ionization gauge (low-pressure).
(*i*) Mercuric oxide tube for the introduction of oxygen.
(*j*) Alkali metal reservoir.
(*k*) Tube being processed.

The system shown is but one of a great many possible systems, only a few of which can be discussed in this treatment. One thing applies to all systems and cannot be emphasized too much; that is, the essential need for cleanliness. The very best of pumps and the most elaborate sys-

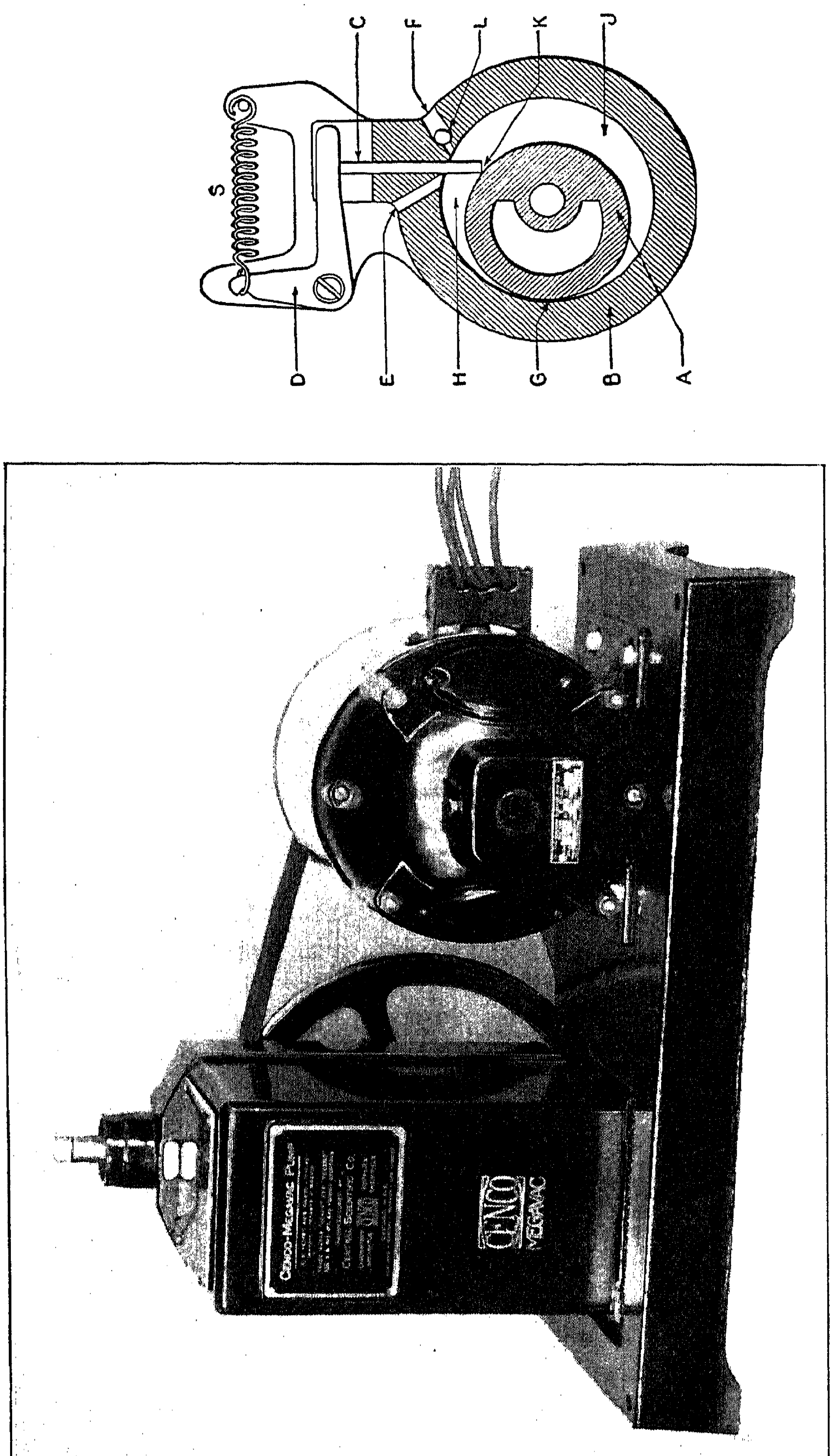


FIG. 5.2.—Cenco Megavac Pump (courtesy of Central Scientific Company).

tem, if contamination is present, cannot possibly yield good results. Because its presence is much less easily detected than leaks or flaws in the vacuum system, contamination due to lack of care and cleanliness is the most common cause of failure in the processing of electronic devices.

5.1. The Fore-Vacuum. All the high-vacuum pumps described in the succeeding section must work into a relatively low pressure. The pressure required in the fore-vacuum depends upon the type of pump exhausting into it and upon the final vacuum to be attained. The range of pressures commonly needed is from 0.1 mm to 0.001 mm Hg.

A number of oil-sealed mechanical pumps are suitable for producing the fore-vacuum. One frequently used type is a rotary oil pump marketed under the name of "Megavac."* The principle of this pump can be seen from Fig. 5.2. The eccentric rotor, A, turning in a counterclockwise direction, compresses the gas in compartment J and forces it out through exhaust vent F. At the same time the volume of chamber H is increasing, allowing gas from the fore-vacuum to enter through intake tube E. The two compartments are separated by the sliding knife C. As the point of greatest radius of the rotor passes the knife edge, the cycle starts over again. The rotor bearings and the valves are immersed in oil, which effectively seals them against gas leakage. In the type of pump named, two of the above-described elements are assembled in a unit to operate in parallel.

It is common practice to connect the intake of the fore-pump to a reservoir of several liters' capacity. The connection is usually made with rubber pressure tubing, or metal piping, and a stopcock is placed between the pump and the reservoir. The reservoir is connected directly to the high-vacuum pump. After the initial exhaust is completed and the high-vacuum pump has reduced the pressure to a low value, the stopcock can be closed and the fore-pump stopped. If the system is tight, the high-vacuum pump can be operated for long periods of time exhausting into the reservoir, since the gas removed by this pump, when raised to fore-vacuum pressure, occupies a very small volume.

5.2. High-Vacuum Pumps. Of the many suitable low-pressure pumps, space will permit only the inclusion of two frequently used varieties, namely, the molecular pump and the diffusion pump. A fore-vacuum must be provided for either of these pumps as they are inoperative at atmospheric pressure.

a. The Molecular Pump. When a gas molecule strikes a surface and rebounds, it leaves in random directions and speeds relative to that surface. Although the rigor of this law is debatable, it certainly holds with sufficient exactness for both a qualitative and quantitative analysis of

* Trademark by Central Scientific Company.

the phenomenon as applied to the molecular pump. If the surface upon which the gas particles strike is moving relative to some reference point, the leaving molecules will have a component of velocity in the direction of motion of the surface. Even though this drift component is small for a single impact, if the process is repeated a large number of times a considerable flow of gas can be produced. Thus, in a chamber with one moving boundary, as shown in Fig. 5.3*a*, a pressure difference can be established between the two ends. As the length of this chamber is increased, the ratio of pressures p_B/p_A increases. This effect is present irrespective of the average pressure in the chamber. When the pressure is high, the pressure difference can be expressed in terms of the viscosity of the gas. As the pressure decreases, collisions between molecules be-

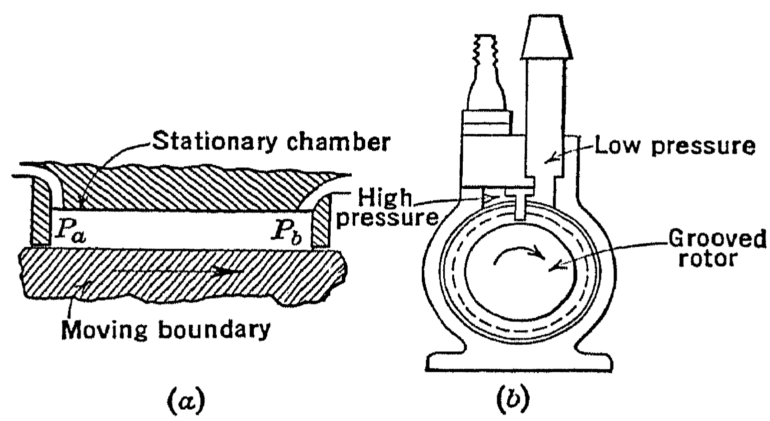


Fig. 5.3.—Principle of the Gaede Molecular Pump.

come fewer in number, and the concept of viscosity loses its meaning, but even where the mean free path is greater than the dimensions of the chamber the effect will be present, since the molecules must strike the moving boundary. Of course, the quantitative relation between the pressure ratio and the velocity, area, etc., of the moving boundary will be different for high and low pressures. This principle was first applied to vacuum pumps by Gaede. In his original pump the moving boundary was formed by a series of circumferential slits in a rotating drum. A set of fixed vanes extended into the slits in such a way that each slit constituted a closed chamber. These were connected in series by suitable exhaust and intake vents.

Applying the same principle, Holweck greatly improved this form of pump. Fig. 5.4 illustrates the pump diagrammatically. The chamber consists of spiral grooves, C and C', extending from the center to the two ends of the pump. The groove decreases in depth from the low-pressure intake at the center toward the exhausts at the ends. The moving boundary consists of a carefully balanced rotor B driven by a small motor

whose armature is enclosed in the pump itself, and is at fore-vacuum pressure. There are no bearings which have to be vacuum tight; there-

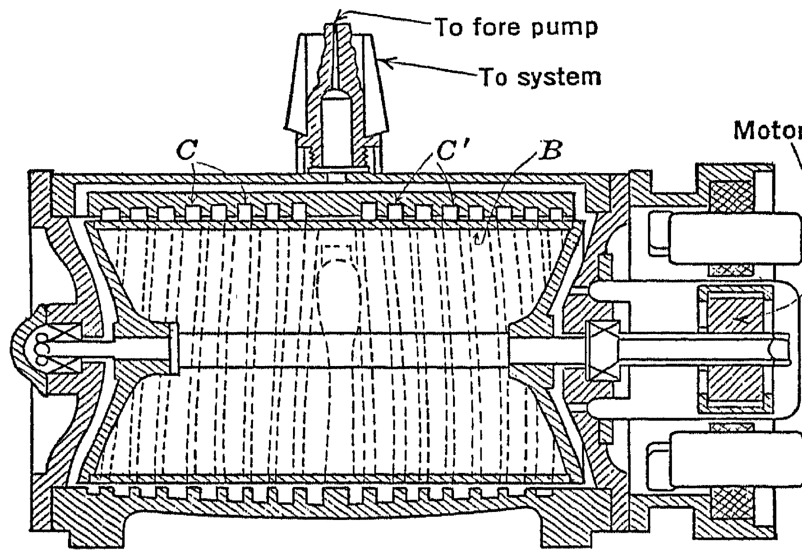


FIG. 5.4.—Holweck Molecular Pump.

fore, the friction can be made very small. When the vacuum is established the rotor can be driven at 4000 rpm with an expenditure of only about 10 watts. The pressure ratio between exhaust and intake can be made to be between 10^6 and 10^8. Thus, with an adequate fore-vacuum to back it up, this type of pump is one of the most powerful tools available for vacuum practice. The pump is rather delicate, however, owing to the necessarily small clearances between the rotor and stator.

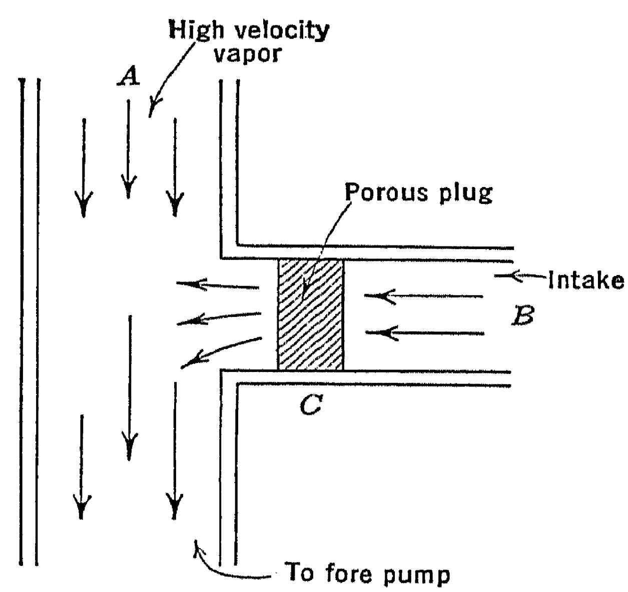


FIG. 5.5.—Principle of Diffusion Pump.

b. The Diffusion Pump. The most widely used high-vacuum pump is the diffusion pump. This pump is extremely effective and rugged, and, because it involves no moving parts (except vapor and gas), it has an almost indefinite life.

The principles involved can be better understood with the aid of Fig. 5.5. In this figure it is assumed that vapor is moving at fairly high velocity through tube A. Tube B is connected to the chamber being exhausted and contains molecules of the gas from this vessel. As molecules of this gas diffuse, due to their thermal motion, out into the stream of vapor in A they will, through collisions with vapor molecules, be given

a drift component down A and thus out of the system. Since the thermal velocity of the vapor molecules is greater than any stream velocity that can be given them in practice, vapor will also diffuse into B. Means must be provided to prevent this diffusion. The porous plug in B serves this purpose. If a suitable vapor is chosen it can also be removed from B by condensation on a suitable cooling jacket. The net action is to remove gas molecules from B without replacing them. Of course, the arrangement as shown would be very inefficient.

Again, it was Gaede who first applied this idea to an actual vacuum pump, employing mercury vapor as the working fluid, and a narrow slit to prevent the backward diffusion of the vapor. As carried out, the pump

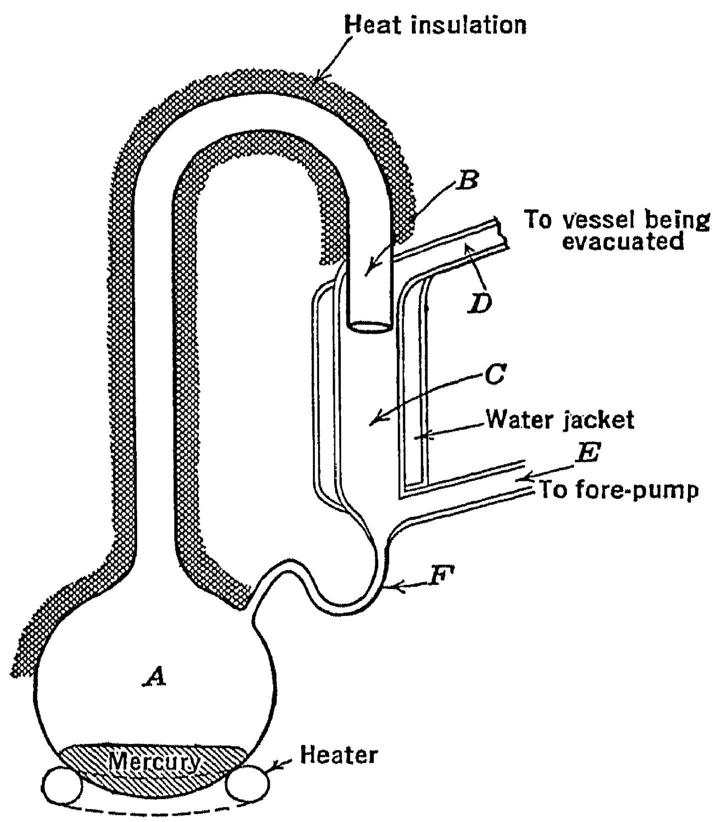


FIG. 5.6.—Diagram of Simple Mercury Diffusion Pump.

was not very efficient and was soon entirely superseded by a pump designed by Langmuir. This pump was based on a similar principle but utilized condensation of the working fluid to prevent backward diffusion. An early form of this pump is shown in Fig. 5.6.

In the pump illustrated, mercury is vaporized in the boiler A, heated by means of an electric heater or a gas flame. The vapor passes up through a heat-insulated tube and out through jet B. Although the mercury vapor pressure is high enough so that there are collisions among the vapor molecules leaving the jet which cause the stream to diffuse to some extent in all directions, nevertheless there is a large average velocity component in the chamber C directed toward the fore-vacuum exit E. Some

of this motion is imparted to the gas diffusing in through the intake *D*. As the mercury vapor comes in contact with the water-cooled walls of the chamber *C*, it condenses and flows back into the boiler through *F*.

A great many modifications of this basic principle have been proposed, some of them resulting in a material improvement in efficiency. Perhaps the most important of these is the use of more than one stage. In this case, the first stage compresses the gas being removed to a pressure midway between that of the high-vacuum side and the fore-vacuum pressure.

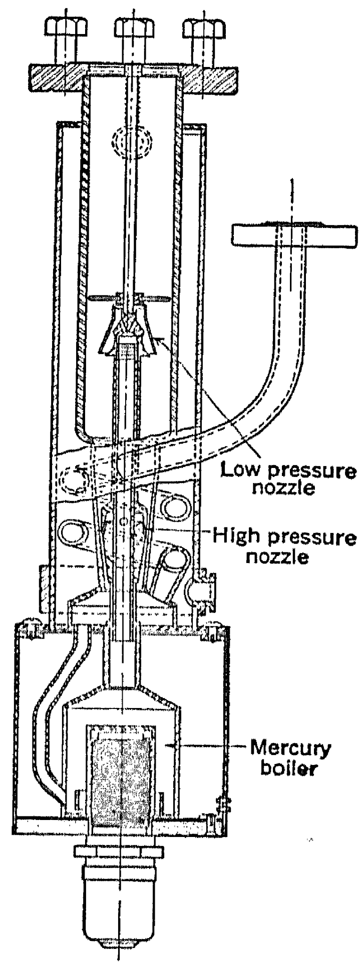


Fig. 5.7.—All-Metal Mercury Condensation Pump (courtesy of General Electric Company).

Although the principle of operation of the two jets is the same, their design must differ if they are to be efficient at their respective pressures. Fig. 5.7 shows a two-stage all-metal pump manufactured by the General Electric Company. The pump illustrated is found to be effective and rugged, and very suitable for general vacuum work.

Diffusion pumps such as described require a liquid-air trap to prevent mercury vapor from going back into the vessel being exhausted, since the vapor pressure of mercury at room temperature—that is, around 30°C —is in the neighborhood of 3×10^{-3} mm.

c. The Oil Diffusion Pump. Certain organic oils are obtainable which

have much lower vapor pressures at room temperatures than mercury, yet which are sufficiently stable to serve as the working fluid in a diffusion pump. Such oils make it possible to dispense with a freezing-out trap.

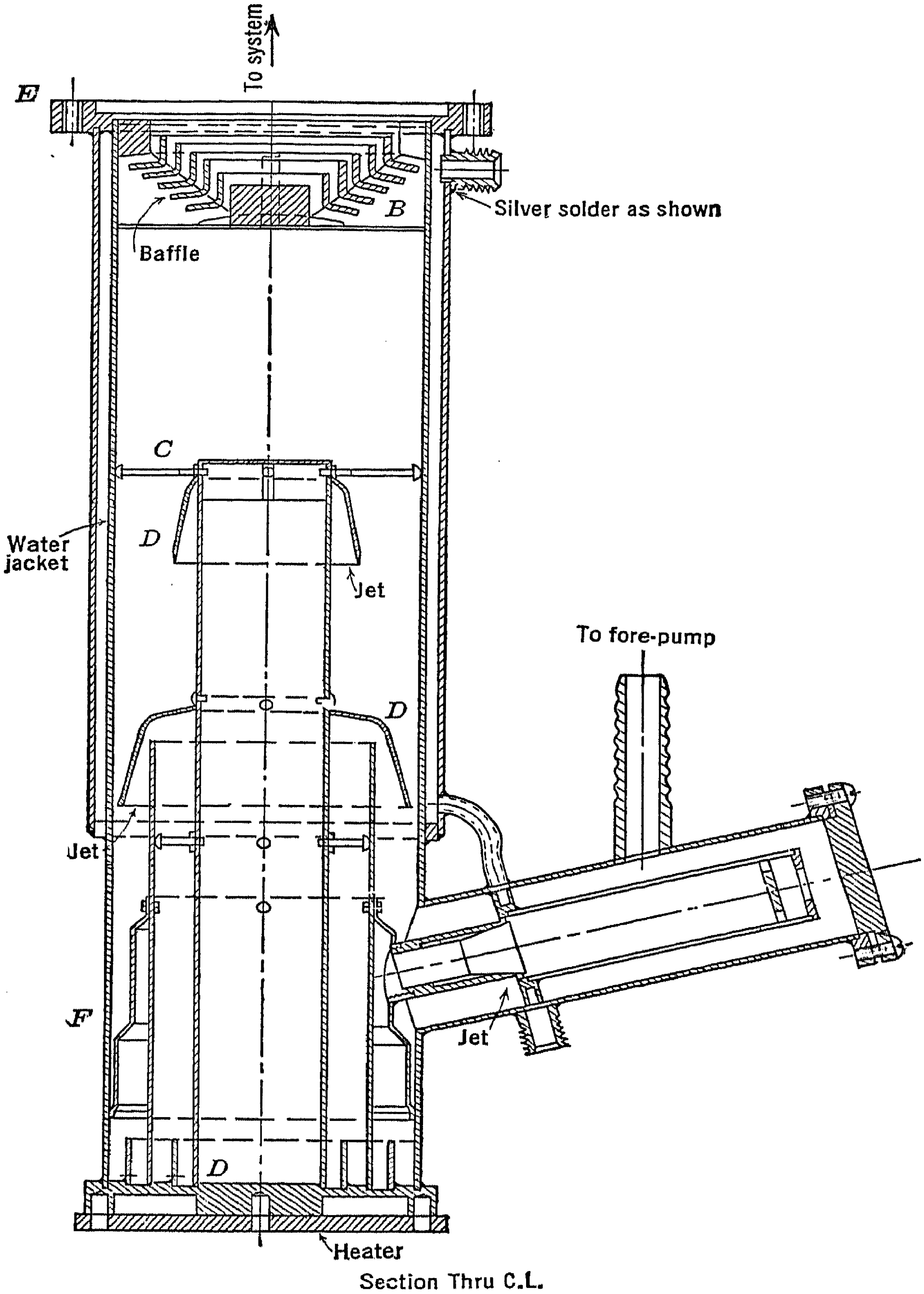


FIG. 5.8.—Oil Diffusion Pump.

Oil diffusion pumps are entirely similar in principle to the mercury pump, and differ only in that they require a slightly altered jet design.

Early workers with this type of pump consistently found that the pres-

sures obtainable were not so low as should be expected from the characteristics of the oil. Hickman succeeded in determining the cause of this apparent anomaly. His experiments indicated that it is due to a fractionating action of the pump which causes the higher-vapor-pressure constituents to appear in fairly large concentration in the working vapor of the pump. As their vapor pressure is below the fore-vacuum pressure these constituents are not removed by the fore-pump, but redissolve in the oil of the diffusion pump and are recirculated. Having isolated the cause, Hickman suggests a number of ways of overcoming this difficulty.

In Fig. 5.8 is shown a successful three-stage oil pump so arranged that the high-vapor-pressure constituents appear only in the high-pressure jet nearest the fore-pump, and never get into the vapor stream supplying the low-pressure jet. With Octoil as a working fluid, this pump will produce pressures of less than 10^{-7} mm Hg without the aid of a freezing-out trap.

5.3. Vacuum Plumbing. The various elements of a vacuum system are interconnected by means of suitable tubing. Consideration must be given to the selection and manipulation of materials for this purpose.

Three types of tubing are common in vacuum practice, namely, glass, metal, and rubber. All these, when special precautions are taken, are suitable for very low-pressure work. However, it is generally accepted that glass or, even better, quartz should be employed where the very best vacuum is required. This is because such vitreous material has an extremely low diffusion rate, can be easily and thoroughly freed from gas and contamination, and the components can readily be joined together.

Brass, copper, and iron tubing are all used in systems even where very low pressures are to be produced. The first difficulty to be overcome is leakage due to the porosity of the metal. Where castings are employed such porosity is unavoidable and the material must be coated with some protective layer. For more or less temporary apparatus, very satisfactory results can be obtained by painting the outside surface with a non-porous lacquer, such as Glyptal. For more permanent equipment, the pores can be sealed by dipping the metal tubing into molten lead or zinc. A heavy layer of evaporated metal on the inside walls is often helpful. Extruded tubing, or worked metal, is much freer from pores than castings and can often be used without any additional precautions. The possibility of leakage through channels and pores should be kept in mind.

Most metals, even those definitely non-porous, are not completely impervious to gases. The diffusion rate depends upon the kind of gas, the particular metal, and the temperature. For ordinary temperatures the wall thickness required to give adequate mechanical strength is great enough so that this type of leakage is not serious. At a few hundred de-

grees centigrade, however, the diffusion of gases through the walls becomes a very real problem.

Joints can be made in metal tubing by welding, silver-soldering, soft-soldering, or with the waxes which are described below. Except for the care required to avoid channels and holes, no particular precautions need be taken. The technique of sealing metal to glass has been developed to such a point that it is perfectly practical for laboratory use. This procedure is discussed in section 5.5.

Rubber tubing finds extensive employment in such high-pressure work as is typified by the fore-vacuum. In general, it is not very satisfactory at very low pressures because most rubber, unless specially prepared, contains high-vapor-pressure components. Furthermore, rubber tubing is difficult to clean and outgas since it cannot be raised to a high temperature.

The walls of rubber tubing used in vacuum work must be thick enough so that they do not collapse under atmospheric pressure. Pressure tubing satisfactory for this purpose has a wall thickness about equal to the diameter of its bore. Joints can be made by slipping the rubber over the glass or metal tube and sealing them with stopcock grease or, where greater permanency is desired, with shellac. Unless the fit is very tight, it is well to wire the joint.

In laying out and constructing a vacuum system, care should be taken to minimize the length of piping used. This is particularly important at low pressures, where short lengths and large diameters are necessary to reduce the resistance to flow of gas in the tubing. Poor design of the system between the low-pressure pump and the vessel being exhausted may completely nullify the advantage of a high-speed pump.

Knudsen* derived the following expression for the flow of gas at low pressure in cylindrical tubing:

$$Q = 3.8 \times 10^3 \sqrt{\frac{T}{M}} \frac{D^3}{L}(P_b - P_a), \tag{5.1}$$

where Q is the volume of gas per unit time (in cm^3/sec) multiplied by the pressure (in dynes/cm^2); D and L are the diameter and length of the tubing; and P_a and P_b are the pressures at the two ends. The temperature in degrees absolute and the molecular weight of the gas are designated by T and M, respectively. This relation can be applied only where the pressure is such that the mean free path of the gas molecules is long compared to the bore of the tube. It will be noticed that the flow varies with the cube of the diameter. Thus, if there is a constriction in the system having a diameter of one-fifth the bore of the tubing used, it will have a resistance to flow equivalent to 125 times its length of tubing.

* See Knudsen, reference 9.

The above relation is useful, for example, for calculating the type of tubing suitable for a given system. To consider a specific case, let it be assumed that a pump has a speed S of 15,000 cm³/sec at the working pressure, the speed being defined as the volume of gas per unit time at the pressure in question that can be withdrawn by the pump. If the connecting tube is to be 30 cm long, and a pumping speed at the vessel of no less than half that of the pump is desired, the diameter of the tubing can be found as follows.

The rate Q for the pump is, of course, $Q = SP$. Since gas flows out of the tube at a rate Q, it must also flow into it at the same rate; therefore, if P_a is the pressure in the vessel being pumped, the pumping speed S_a at the vessel will be given by:

$$S_a P_a = Q. \tag{5.2}$$

Combining these equations with Eq. 5.1, the following is obtained:

$$\left(\frac{1}{S_a} - \frac{1}{S}\right) 3.8 \times 10^3 \sqrt{\frac{T}{M}} \frac{D^3}{L} = 1. \tag{5.3}$$

Evaluating Eq. 5.3 from the data given, the diameter needed is:

$$D = 4.19 \text{ cm.}$$

It should be noticed that the working pressure P_a will be twice the pressure at the pump.

For the parts of the system where the pressure is high there no longer exists the need for large-bore, short-length tubing, and, in general, diameters of ¼ inch or so can be used. A system can, therefore, be laid out to minimize the length of the high-vacuum parts without particular consideration of the resulting length of the fore-vacuum side.

Cutoffs and stopcocks are often necessary as a part of a vacuum system. In general, stopcocks are used only where gas pressures are high, although there are numerous exceptions. Two reasons for this are the difficulty in preventing some leakage over large ground joints and the serious resistance to the flow of gas at low pressures presented by the small bore of the usual stopcock. Mercury cutoffs, however, are satisfactory, provided that there is a freezing-out trap between the cutoff and the vessel being exhausted. Three typical cutoffs are illustrated in Fig. 5.9. That shown in Fig. 5.9*a* has a glass cap which can be lowered over the exhaust vent and sealed with mercury. The motion is accomplished by means of an iron armature and an external magnet. The second type (Fig. 5.9*b*) is closed by raising the level of a mercury column until it seals a large-bore U-tube.

Stopcocks are occasionally used on the high-vacuum side of a system to admit small amounts of gas when needed for processing the tube being

evacuated. Such low-pressure stopcocks are often sealed with mercury as well as stopcock grease. Furthermore, special greases having low vapor pressures and decomposition rates are necessary under these conditions. However, since a number of greases available on the market meet this requirement, this does not present a serious difficulty.

A final type of cutoff, suitable for admitting minute quantities of a gas into the system, should be mentioned. This makes use of the diffusion of gas through porous material and is shown in Fig. 5.9*c*. The chamber E is divided into two parts by partition F, the intake and exit tubes entering these at the bottom as shown. The porous plugs A and B are sealed

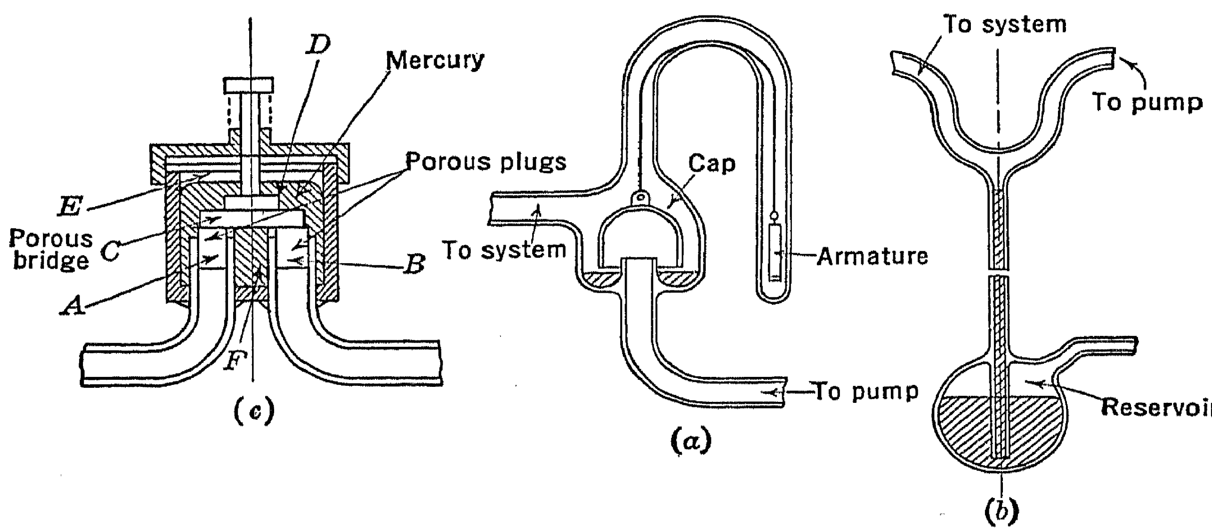


FIG. 5.9.—Typical Vacuum Cutoffs.

into the entrance and exit divisions. A small amount of mercury D covers the top of the two plugs. The member C is also porous and is mounted so that it can be lowered into contact with A and B. When it is in contact the gas can diffuse from A into C and thence into B and the system. When C is raised both the intake and exit are sealed by the mercury and no gas can flow.

5.4. The Measurement of Low Pressures. The measurement of the pressure in a vacuum system is important both from a theoretical and a practical standpoint. With a knowledge of the pressure, the number of gas molecules present and the rate at which they impinge upon a given surface can be calculated. In practical exhaust technique, it is desirable to be able to measure pressure, as it is an index of the cleanliness of the system, of the completeness of the outgassing, and of the presence of leaks. It is also necessary to measure the pressure when gas dosage is part of the activation schedule.

Unfortunately, no single gauge is suitable for the entire range of pressures encountered in the vacuum practice required in electronics.

Four types of gauges, covering different pressure ranges, will be de-

scribed. These are the manometer, the McLeod gauge, the thermocouple gauge, and the ionization gauge. This list is not exhaustive, of course, and other measuring equipment is preferred by some operators, but those described are simple, rugged, practical, and yet sufficiently accurate for nearly all measurements encountered in tube work.

The manometer is perhaps the most direct pressure-reading device; i.e., there is a minimum of theory between the gas pressure and the number read on the gauge. However, it is not sufficiently sensitive to be practical for pressures much below 10^{-1} mm. Fig. 5.10*a* illustrates the typical manometer. The U-tube is closed at the end C and connected to the system at D. A liquid, B, having low vapor pressure and a known

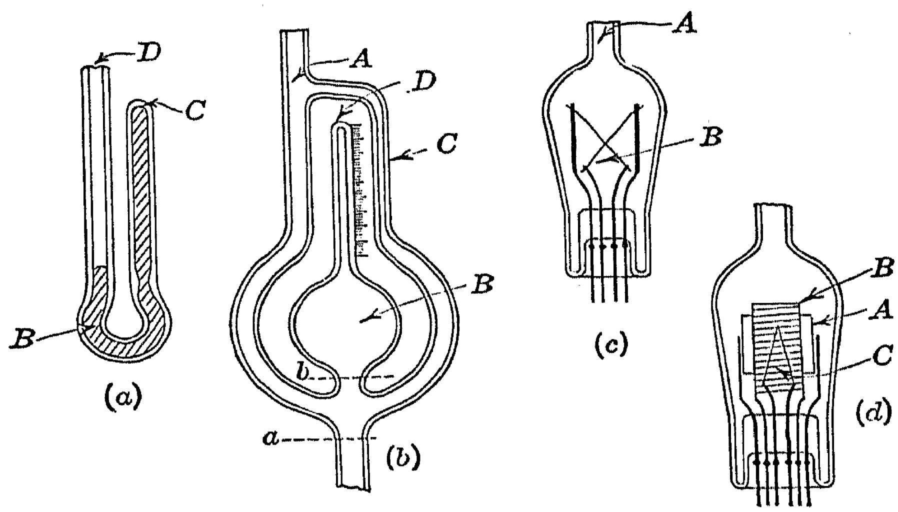


FIG. 5.10.—Vacuum Gauges.

density ρ, partly fills the U-tube. If the pressure on the sealed side is p_0 and that on the system side is p_x, the difference in the heights of the two columns Δh will be $(p_x - p_0)/\rho$. Usually, the arm C is completely filled with liquid when arm D is at atmospheric pressure and the liquid is thoroughly freed from gas. The pressure p_0 is consequently negligible compared with any pressure that can be read with the gauge, making the reading of the gauge simply:

$$p_x = \rho\Delta h.$$

Mercury is often used in this type of gauge. In that case, the pressure can be read directly in millimeters of mercury. Greater sensitivity can be attained with a liquid of lower density, such as *n*-dibutylphthalate, or Octoil. With such a liquid the pressure in millimeters of mercury will be the difference in height multiplied by the ratio of the density of the liquid used to that of mercury. Still greater sensitivity is possible with comparators for measuring the difference in height or with various other devices, but this is rarely worth while in view of the simplicity of the other gauges which operate at lower pressures.

One of the most widely used vacuum gauges is the McLeod gauge. This measuring device takes a specific quantity of the gas in the system, compresses it into a small fraction of its original volume, and, by means of a manometer arrangement, measures the pressure of the compressed gas.

The means and arrangement for carrying this out are shown in Fig. 5.10*b*. The tubing *A* is connected to the system whose pressure is to be measured, and also to the bottom of the chamber *B*. Joined to this chamber is a tube leading to a mercury reservoir. The capillary tube *D*, closed at the top, is affixed to the upper part of *B*. When the mercury is at the level indicated by *a*, the capillary and bulb *B* have free access to the system and the pressure is equal throughout. As the mercury level is raised to *b*, the bulb is trapped off from the system. Further raising of the mercury level compresses the gas remaining in the bulb. To measure the pressure in the system, the level of the mercury is raised until it coincides with a predetermined index in capillary *D*. The gas pressure p_c in the capillary is now equal to the pressure in the system p_0 multiplied by the ratio of the volumes of bulb to capillary. Thus, if the difference in the height of the mercury columns in *A* and *D* is Δh, the pressure in the system will be:

$$p_0 = \frac{V_D}{V_B}\Delta h.$$

The ratio of volumes, which is determined by the gauge construction, may be very great, making this type of gauge a very sensitive measuring device.

Instead of mercury heights being read directly on column *A*, the capillary loop *C* is attached to *A* and readings are made on it. Since this capillary has the same bore as *D* it obviates the necessity of correcting for capillary effects.

A second method of using this gauge is to raise the mercury to some predetermined height in column *A*. If, now, the cross-sectional area of the capillary is *s*, and the distance between the index and the top of the capillary is *l*, then the pressure in the system will be:

$$p = \frac{(\Delta h - l)s\Delta h}{V_B}.$$

When the index coincides with the top of the capillary, this becomes:

$$p = \frac{s\Delta h^2}{V_B},$$

where s/V_B is known from the construction of the instrument.

The McLeod gauge, though simple, sensitive, and accurate, has two disadvantages. It cannot be used for making continuous readings of pressure, and if there are condensable vapors in the system it will give a false reading. This false reading is due to the fact that the operation of the instrument is based upon the applicability of Boyle's law to the gas being measured, and if a substance is present having a second phase whose vapor pressure lies between the pressure in the system and that in the capillary, this substance changes its phase when compressed in capillary D, and does not follow Boyle's law.

A very convenient gauge, which covers the same range of pressures as that covered by the McLeod, is the thermocouple gauge. This type gives continuous readings, and reads pressures of condensable gases. However, it must first be calibrated against known pressures.

The gauge consists of a thermojunction of low heat capacity welded to a thin heater wire. A constant current is supplied to the heater, and, since most of the heat lost is by conduction through the gas around the heater, its temperature is a function of the pressure in the system to which it is attached. Fig. 5.10*c* shows the construction of this type of gauge. The bulb is joined to the system through A. The heater B in the gauge illustrated is of 5 mil tungsten wire; the thermojunction is chromel alumel, or copper constantan, of about the same diameter. The maximum temperature reached by the heater when the pressure is low is around 300°, and the current required to give this temperature is around 100 milliamperes. No advantage is gained by operating the gauge at higher temperatures as radiation losses mask changes in heat transfer by conduction. This type of gauge is suitable over the range from several millimeters to about 10^{-4} mm Hg.

An ionization gauge is the most practical instrument for measurements of really fine vacua. At pressures below 10^{-3} mm Hg the reading is closely proportional to the pressure, and as yet no lower limit for the gauge has been reached. The action of the instrument is as follows: In Fig. 5.10*d* the grid B is made, for example, 100 volts positive with respect to the thermionic cathode C, while the plate A is 20 volts negative with respect to C. Electrons emitted from C are drawn toward and pass through grid B into the region between B and A. They of course eventually return to and are collected by B. If there are gas molecules in the space between A and B, the electrons cause a certain fraction of these to become ionized. The positive ions thus produced are collected by the negative plate A. The positive-ion current I_2, which is measured by a microammeter, is proportional to the electron current I_1, emitted by C, and to the number of ions in the region between A and B. Thus, for a given value of I_1 and

a specific gas, I_2 is proportional to the pressure (for low pressures). Such a gauge must be calibrated for each kind of gas to be measured.

A practical gauge of this type can be made from the parts of an RCA 210. The parts are assembled on a Pyrex press in the usual way and sealed into a bulb tubulated at the top. The gauge is attached by this tubulation

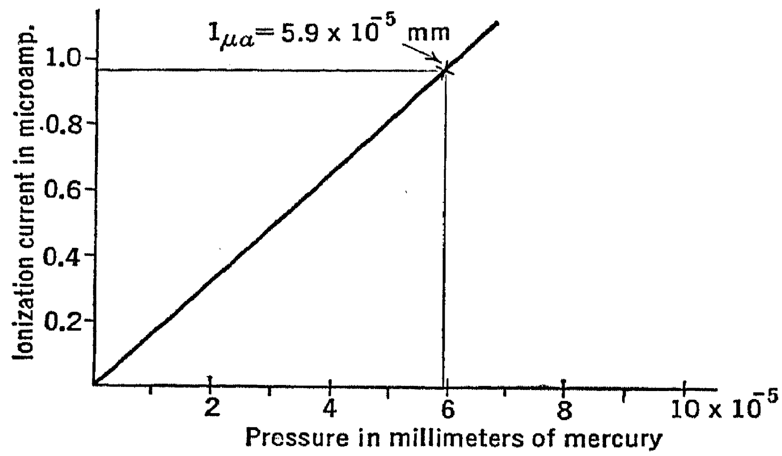


Fig. 5.11.—Calibration Curve of Ionization Gauge Illustrated in Fig. 5.10*d*.

to the system. Fig. 5.11 shows the calibration curve for a particular gauge of this type for air.

Since most hard electronic tubes should be pumped to a pressure such that the ion current lies between 10^{-1} and 10^{-2} microampere, a rather sensitive meter is required for its measurement. A simple, inexpensive and reliable vacuum-tube meter suitable for this purpose can be built

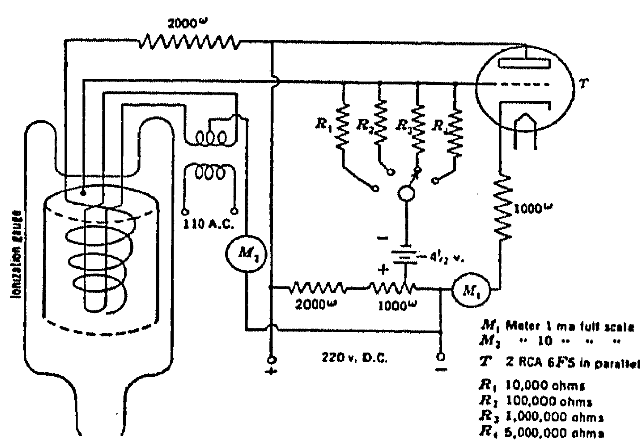


Fig. 5.12.—Circuit Diagram of Vacuum Tube Microammeter Suitable for Use with Ionization Gauge.

according to the circuit diagram in Fig. 5.12. It should be noted that this meter is of the negative feedback type, so that its calibration is independent of the tube characteristics. Furthermore, the output meter can be of the inexpensive panel type.

The four gauges described in this section should meet all practical needs of the experimenter in the field of electronics.

5.5. Vacuum-Tube Construction. Foremost in a consideration of the construction of vacuum tubes is the question of the material forming the envelope. For general experimental work, glass is found to be the most satisfactory. Glass can easily be formed into any desired shape by an experienced glassblower. The material is almost completely impervious to gas so that a tube, once exhausted, retains its vacuum almost indefinitely. Other valuable properties are its high electrical resistance, ability to withstand temperatures of the order of 500°C, chemical inactivity, and transparency.

Various kinds of glasses differ in physical properties. The type of glass selected, of course, depends upon the conditions which have to be met by the tube under construction. A few of the more common glasses, together with their physical characteristics, are listed in Table 5.1.

TABLE 5.1

PHYSICAL PROPERTIES OF GLASSES

Type	Softening Point	Annealing Point	Coefficient of Expansion	Resistance Megohms/cm³	Density	Index of Refraction
Lead	630	431	87×10^{-7}	11,900	3.04	1.557
Lime (soft)	696	510	92×10^{-7}	2.23	2.47	1.512
Fernico or Kovar seal glass	697	484	46×10^{-7}	5300	2.24	1.480
Nonex (hard)	756	521	36×10^{-7}	1170	2.35	1.487
Pyrex (hard)	818	553	32×10^{-7}	263	2.23	1.474

Perhaps the glass most used for experimental tubes is of the type represented by Pyrex. This is because its low coefficient of expansion makes it very resistant to thermal shock. Furthermore, its high melting point allows a high-temperature bake. It should be pointed out that these properties do not obviate the necessity of annealing the glass after it has been heated to a high temperature. If this is not done there is danger of the glass cracking from internal strain if scratched, heated, or subjected to a slight mechanical shock.

So-called soft glasses are also often used because their low melting points make them easy to work. However, much greater care must be taken in the annealing of the glass. Also, the finished article cannot be subjected to as severe thermal shock as Pyrex, because no known soft

glass has so low a coefficient of thermal expansion. The handling of glass is an art in itself, and space does not here permit even a limited discussion of glassblowing, except for the mention of a few special techniques required in electronics problems.

Almost every tube used for either experimental or practical purposes employs conductors passing through the glass walls. These leads may be in the form either of individual wires, or of a press carrying a number of wires. With both constructions there must be a vacuum-tight junction between the metal and the glass itself. The metal used in these seals depends upon the glass, for not only must the glass "wet" the metal, but also the coefficients of expansion of the two materials must match. For lead or lime glass the two most commonly employed materials are plati-

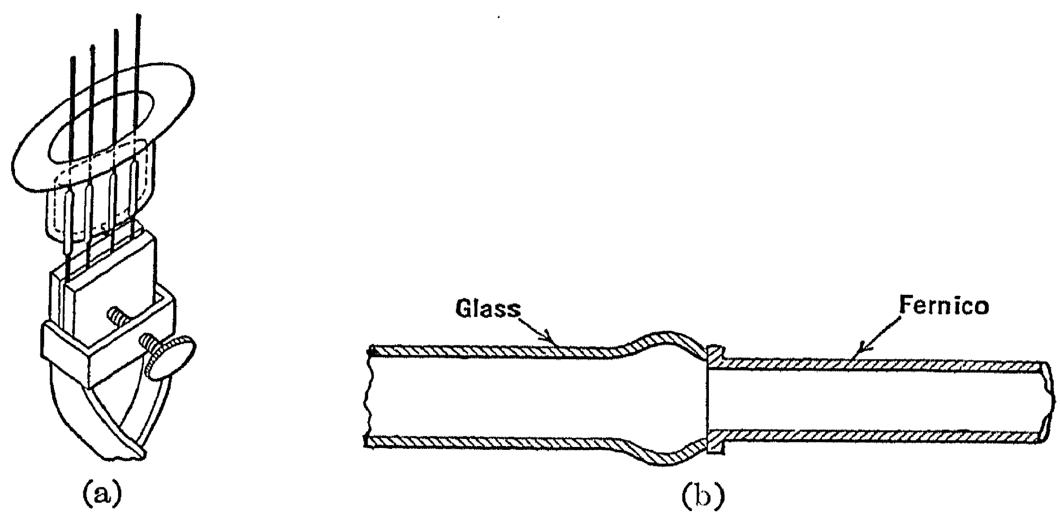


Fig. 5.13.—Permanent Glass-to-Metal Seals.

num and Dumet. The latter is a copper-clad, nickel-copper alloy having an expansion coefficient of about 9×10^{-6} per degree Centigrade, and is much cheaper than platinum. For Pyrex, tungsten is usually satisfactory. Because of its mechanical properties, tungsten is not very suitable for the leads themselves, and it is customary to weld nickel, molybdenum, or copper wires to the short length of tungsten which actually goes through the glass.

To prepare the tungsten for the seal it must first be polished to remove longitudinal die marks, then cleaned with an etching agent like hot potassium nitrite, and finally dipped in distilled water to remove the etching agent. To make the seal, the tungsten is oxidized by heating in air and then beaded with a small droplet of Pyrex. The color of the oxidized tungsten under the bead should be a uniform golden yellow. A gray oxide color, or the presence of bubbles on the surface, often means that the seal will leak. The bead can be sealed into the wall of the blank, or into a press, as shown in Fig. 5.13*a*.

The glass known as Nonex can be beaded onto tungsten in the same way. While it is somewhat easier to make a Nonex seal, the properties of this glass are not so desirable for purposes of making a tube as those of Pyrex. Often, presses are made of Nonex and sealed into a Pyrex blank with an intermediate band of Uranium glass.

It is sometimes necessary to seal fairly large pieces of metal to glass. Copper can be successfully sealed to either hard or soft glass, using a technique attributed to Houskeeper.* In this process no attempt is made to match the expansion of glass and metal, but instead the copper is tapered to paper thinness and the thin copper is sealed to the glass. The thin copper is sufficiently ductile to yield to the expansion or contraction of the glass.

One of the most easily made metal-to-glass seals uses Fernico, Kovar, or a similar iron-alloy metal and a special seal glass. The metal seal can, of course, be welded or soldered to the metal part being joined, while the glass end of the seal can be affixed to the rest of the system by means of a graded seal. The actual joining of glass to the alloy is simplicity in itself. It is merely necessary to clean the metal where it is to be sealed, form a thin oxide layer by heating in air, and then press the heated glass and metal together. It is not necessary to flow the glass over a large area; a joint little thicker than the walls of the tubing being joined usually suffices. Fig. 5.13*b* illustrates a typical joint.

The use of metal envelopes for electron tubes is rapidly becoming more extensive. This technique involves a combination of machine work and glassblowing. The envelope is spun or pressed into shape, usually in two parts. Holes are drilled for the leads. These leads are usually of alloy, beaded with seal glass, so that the beads can be joined to the edges of the holes. The two halves of the shell are joined by brazing with molten copper. The completed tube is pumped through a short length of metal tubing which can be crimped together to seal it off the system and then permanently closed with solder.

Methods for making metal-to-glass seals with a number of iron alloys have been perfected. This is largely due to extensive research on the part of the glass industries which have made available glasses whose coefficients of expansion match those of the alloys.

Ceramic envelopes, and ceramic-to-metal seals, are also beginning to enter the field of vacuum practice. Although certain advantages are to be derived from the use of these materials in industrial operations, their manipulation is not sufficiently simple to make them practical for general experimental purposes.

The worker has at his disposal a wide variety of materials for the

* See reference 10.

internal construction of electronic devices. The only general restriction is that of vapor pressure. Except for this, the choice is dictated by the special requirements of the problem.

As insulation, mica, Lavite, porcelain, and glass are most useful. All these materials can stand a high temperature, have a low vapor pressure, and are easily worked.

Nickel is a very practical metal for vacuum work. Its mechanical properties are good, the melting point is high, and it is fairly inert chemically. Occasionally, the moderately high magnetic permeability makes it undesirable. Aluminum, copper, nichrome, Monel, and numerous other metals or alloys can take its place.

Where high temperatures are to be encountered, either in the processing or operation of the device, tantalum, molybdenum, platinum, or tungsten is suitable.

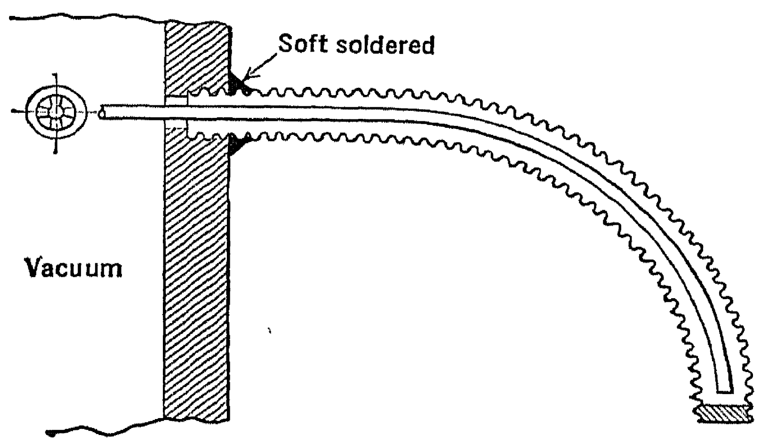


FIG. 5.14.—Crank Enclosed in Flexible Sylphon Tubing.

It is often desirable to introduce mechanical motion into experimental electron tubes. The simplest method is perhaps the utilization of gravity to move a sliding member or to rotate an unbalanced wheel. Another very simple expedient is to move or rotate a suitable armature by means of an external magnetic field. Various electrical expedients will suggest themselves to the reader, such as an electrically heated bimetallic strip. A very useful, purely mechanical device is a curved crank and some sylphon tubing. This is illustrated in Fig. 5.14.

Probably the most useful tool for vacuum construction is the spot welder. A relatively simple condenser-powered spot welder is illustrated in Fig. 5.15. The circuit of the power supply is shown in Fig. 5.16. The device can, of course, be altered to meet the immediate requirements of the operator, but some form of welder is almost an essential.

5.6. Exhaust and Processing of Vacuum Tubes. After the tube to be activated has been attached to the system and every precaution has

been taken to insure that there are no air leaks, it must be thoroughly outgassed.

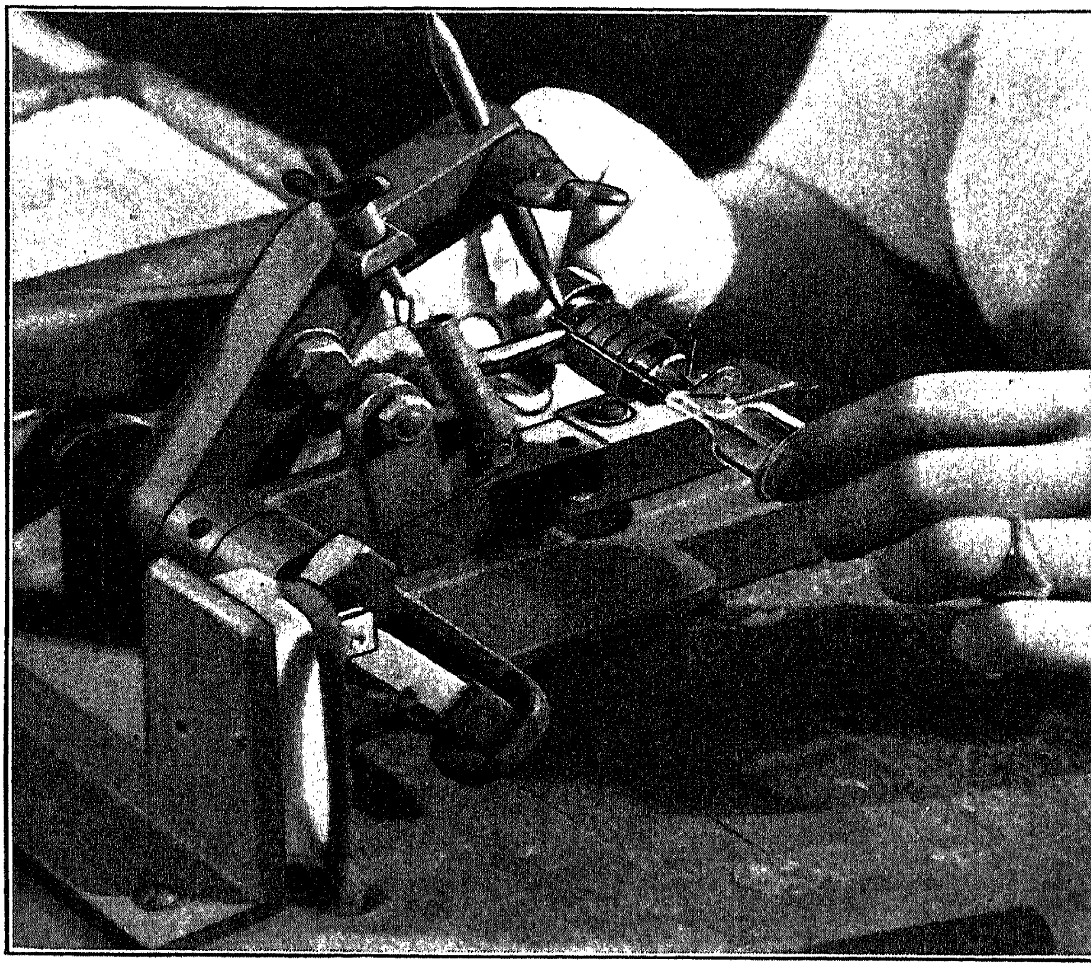


Fig. 5.15.—Spot Welder.

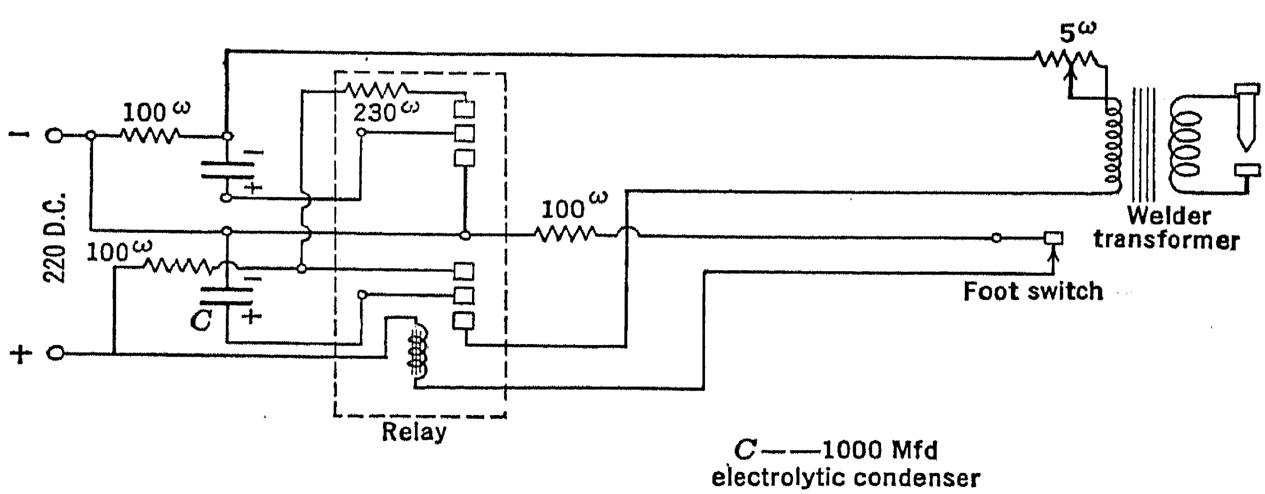


Fig. 5.16.—Circuit Diagram of Impulse Welder.

The outgassing is accomplished by baking the tube at as high a temperature as is permitted by the glass envelope and the materials in the tube.

A Pyrex envelope can be baked at 450°C; a soft glass vessel should not be heated above 350°C. In order to avoid fouling the freezing trap with the large quantities of condensable gas that are emitted by the tube

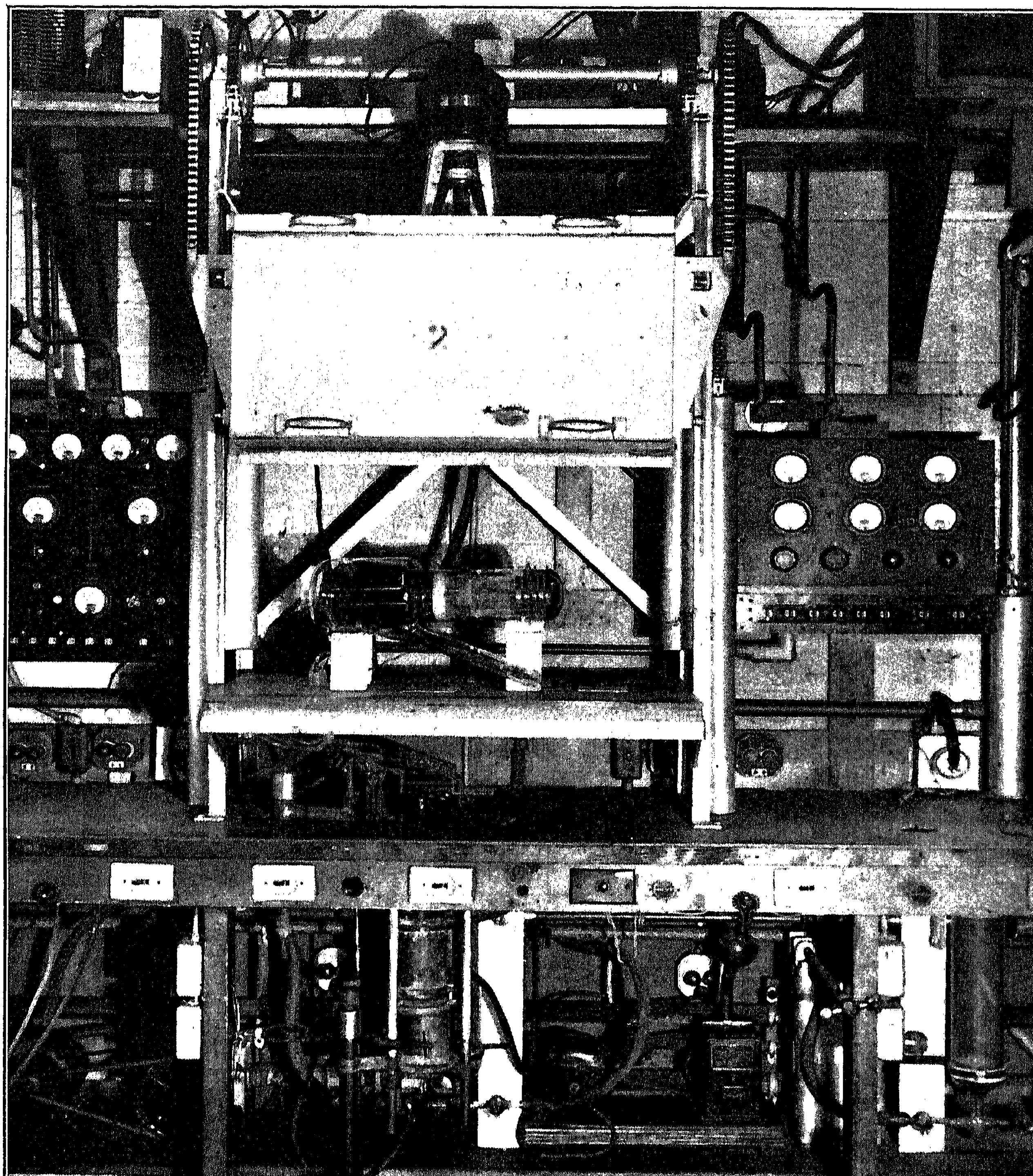


FIG. 5.17.—Vacuum System Used in Electronic Research.

when it is first heated, the bake should be started with both the fore-vacuum and diffusion pump in operation but without the trap. After the tube has reached baking temperature, the freezing trap is used. The bake is usually continued until the pressure in the tube drops below 10^{-5}

mm Hg. Where there is much metal in the tube it is desirable, when possible, to heat the metal to a high temperature with a high-frequency bombarder. Fig. 5.17 is a photograph of a typical system showing the baking oven.

The vacuum in a sealed-off tube can be very much improved by means of a getter. In fact, present high-speed production methods of commercial vacuum-tube manufacture would be impossible without getters.

The getter is an active substance which is capable of taking up and binding a large quantity of gas. It is usually released inside the tube just before it is sealed off the system, and it continues to collect gas for a long period of time. The action of a getter may be purely chemical, it may be by adsorption, or the gas may dissolve in the getter. Most getters function in two or more of these ways.

Phosphorus was probably the first material used as a getter. This element is extremely active in cleaning up residual oxygen. Barium, calcium, and magnesium are all excellent getters, their abilities as clean-up agents decreasing in the order of their listing. However, because of the difficulty of manipulation, calcium and magnesium preceded barium in general use. Recently, barium sealed into small copper tubes, or alloyed with tantalum, has become available on the market, which has greatly increased the practicability of this material as a getter. An alloy of the rare-earth metals, under the name Mischmetal, is also extensively used.

The getter is generally applied in the form of a thin evaporated film covering portions of the walls of the tube. The film is usually obtained from a small amount of the getter material attached to a disk of refractory metal which is heated with a high-frequency bombarder. In addition to facilitating the heating of the getter, the disk acts as a shield to prevent the getter from depositing onto parts of the tube where it may produce harmful effects.

The processing of the tube depends entirely upon the use to which it is to be put. A great many tubes encountered in television work involve photosensitive surfaces activated with caesium on silver oxide. Mention of the special equipment necessary for this treatment is therefore not out of place. The silver is oxidized by glowing it with an electrical discharge in oxygen at a low pressure. Oxygen can be conveniently admitted to the system from a side tube containing red mercury oxide which, upon heating, gives off pure oxygen.

Caesium can be very easily introduced by the reduction of combined caesium. Often the caesium compound, together with a reducing agent, is pressed into pellets designed to release a definite amount of caesium upon heating. One of the several compounds which can be used is caesium chromate with tantalum powder as reducing agent. Azide compounds,

which break up into the alkali metals and nitrogen when heated, are also frequently used. These pellets are mounted in metal cups, which can be heated with a high-frequency bombarder. The pellets and holders are usually placed in side tubes attached directly to the bulb being processed rather than to the system.

The other alkali metals can be introduced in a similar fashion.

Finally, some mention should be made of the technique of evaporating metals and insulators. The evaporation may be done in the final tube, or in a demountable system. Unless a very pure, clean metal surface is desired, an extremely low pressure is not required. However, the vacuum should be good enough so that the mean free path of the molecules is considerably greater than the distance between the source and the surface upon which the film is being deposited. A very convenient way of making the source is to place the material to be evaporated in a helical spiral of tungsten wire. An alternative method, when a ductile metal is to be used, is to place a number of small loops of the metal on a tungsten filament. These loops melt into small drops which adhere to the filament by surface tension. These are only two of a great many possible sources, and, in general, the type of source depends upon the particular problem.

The surface upon which the metal is to be deposited must be very clean or else the metal will not stick to the surface. Oils and greases should be removed with an organic solvent and then the surface cleaned with an acid. When possible, additional cleansing can be effected by glowing the surface with a gas discharge.

5.7. Demountable Vacuum Systems. Demountable systems are of considerable importance in experimental electronics. It is possible to use such systems not only for purposes of evaporating metal or insulating films, vacuum cleaning, and similar routine problems, but also for investigating test structures involving thermionic or photoelectric cathodes. Commercially, demountable systems are important in connection with short-wave transmitting tubes, X-ray tubes, etc.

In general, a demountable system necessitates at least one large sealed joint. Where the system is semi-permanent and does not have to be opened except at infrequent intervals, the best seal is made with a soft metal ring clamped between two ground metal flanges. Semi-permanent joints between metal and glass, porcelain, or ceramic materials can also be made by coating the latter with a thin platinum film and soldering the metal directly to the platinum. The metal film can be applied to the glass or ceramic surface in the form of a platinum chloride suspension in essential oils,* which reduces to metallic platinum upon heating. A joint

* A commercial preparation suitable for this purpose is sold under the name of Hanovia Platinum Bright.

of this type is shown in Fig. 5.18*c*. This method is not very practical if the joint has to be broken frequently and it is often more convenient to use an organic wax. Waxes for this purpose may be arbitrarily divided into four groups, namely: hard, high-temperature waxes; flexible, moderate-temperature waxes; soft, low-temperature preparations; and compounds which can be applied at room temperature.

A description will be given of a representative member of each category, but it should be understood that actually there are a great many waxes in each group.

A useful hard wax can be made from white shellac and beeswax mixed in proportions from 1:1 to 9:1, depending upon the rigidity required, and melted together to form a homogeneous mass. This wax melts at

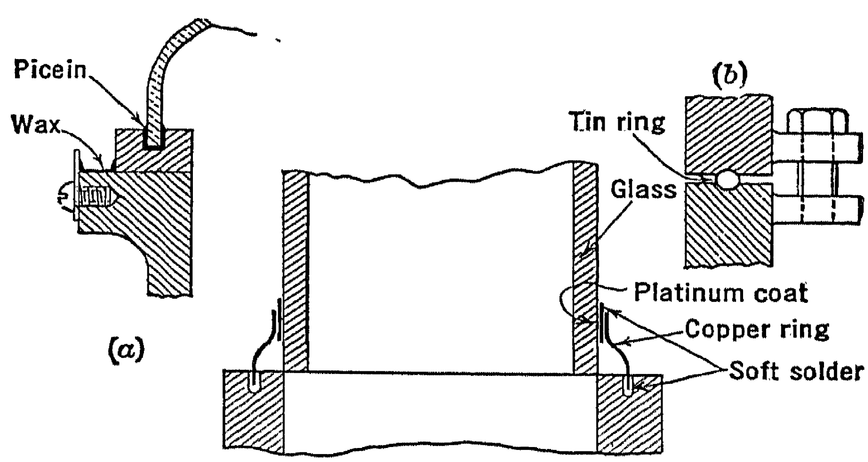


FIG. 5.18.—Vacuum Seals for Demountable Systems.

about the temperature of boiling water. In applying the wax it is necessary to heat the surfaces being joined as well as the wax itself. If this is not done the wax will not "wet" the surfaces and the joint will not be vacuum tight. This type of wax is hard enough so that a considerable area will withstand atmospheric pressure, and joints sealed with the material do not need to be close fitting. Furthermore, the compound has sufficient mechanical strength to stand considerable load.

Picein is a much more flexible wax than that mentioned above. It melts at about 80°C and has a vapor pressure of approximately 5×10^{-4} mm Hg at room temperature. As is true of the hard wax, it is necessary to heat the two surfaces being joined if the seal is to be vacuum tight. This wax is firm enough to seal joints which are not close fitting, but it will flow under pressure and cannot be used to close large openings or where a member is subjected to much mechanical strain. Its flexibility makes it useful where there is much vibration, since it does not crack or fracture like a harder wax.

For a soft, low-temperature wax, beeswax is often used. In order to remove the more volatile compounds present in the natural wax, the material should be heated at 100°C for several hours under low pressure. Joints made with this wax should be between close-fitting surfaces, i.e., surfaces which are flat but not necessarily ground. The wax is not applied to the surfaces directly but to the edges of the joint as illustrated in Fig. 5.18*a*. It is not necessary to heat the surfaces being joined.

Certain compounds are available which require no heat for their application. The preparation going under the trade name of Apiezon Q is an example of such a material. At room temperature it has about the consistency of ordinary modeling clay. It is applied in much the same way as ordinary putty. For temporary joints which are not raised far above room temperature and which are not subject to any mechanical strain, this type of wax provides a very quick and convenient way of making vacuum joints.

Before leaving the subject of demountable systems, the soft metal seal mentioned above should be described in greater detail. The two surfaces to be sealed should be smooth, preferably ground. A narrow groove, not more than 1/32 inch in depth, is cut in one surface. The sealing metal is made into the form of a wire ring of the same diameter as the groove. Ordinary type metal is a very suitable material for this ring. The ring is placed in the groove, and the two surfaces are clamped together with sufficient pressure to cause the sealing metal to flow. Such a joint is illustrated in Fig. 5.18*b*. When well made, this seal can be used in systems requiring pressure lower than 10^{-5} or 10^{-6} mm Hg.

There are a great many tricks of technique in connection with work with demountable systems, such as the use of Duco cement, Cellophane scotch tape, Glyptal, and Canada balsam. However, space does not permit their inclusion in this very brief account.

For the more complete information which is required for any extensive experimenting in the field of electronics, the reader is referred to the excellent discussions given in the bibliography below.

REFERENCES

1. Saul Dushman, "Production and Measurement of High Vacuum," *Gen. Elec. Rev.*, Schenectady, N. Y., 1922.
2. F. H. Newman, "Production and Measurement of Low Pressures," Van Nostrand Company, New York, 1925.
3. L. Dunoyer, "Vacuum Practice," G. Bell and Sons, Ltd., London, 1926.
4. M. Knoll, F. Ollendorff, and R. Rompe, "Gas Discharge Tables," Julius Springer, Berlin, 1935.
5. W. Espe and M. Knoll, "Materials of High Vacuum Technique," Julius Springer, Berlin, 1936.

6. Saul Dushman, "Recent Advances in the Production and Measurement of High Vacua," *J. Franklin Inst.*, Vol. 211, pp. 689–750, 1931.
7. W. Gaede, "Molecular Pump," *Physik. Z.*, Vol. 13, pp. 864–870, 1912; *Ann. Physik.*, Vol. 41, pp. 337–380, 1913.
8. The Vapor Pump:

W. Gaede, *Ann. Physik.*, Vol. 46, pp. 357–392, 1915.

I. Langmuir, *Gen. Elec. Rev.*, pp. 1060–1071, 1916; *J. Franklin Inst.*, Vol. 182, pp. 719–743, 1916; *Phys. Rev.*, Vol. 8, pp. 48–57, 1916.

C. R. Burch, *Nature*, Vol. 122, p. 729, 1928.

K. C. D. Hickman and C. R. Sanford, *Rev. Sci. Instruments*, Vol. 1, pp. 140–163, 1930.

K. C. D. Hickman, *J. Franklin Inst.*, Vol. 213, pp. 119–154, 1932; Vol. 221, pp. 215–235, 1936; Vol. 221, pp. 385–402, 1936.

L. Malter and N. Marcuvitz, *Rev. Sci. Instruments*, Vol. 9, pp. 92–95, 1938.

9. M. Knudsen, "Flow of Gases," *Ann. Physik*, Vol. 28, pp. 75–130, 1908; Vol. 28, pp. 999–1016, 1908; Vol. 44, pp. 525–536, 1914.
10. Glass-Metal Seals:

W. G. Houskeeper, *J. A. I. E. E.*, Vol. 42, pp. 954–960, 1923.

A. W. Hull and E. E. Burger, *Physics*, Vol. 5, pp. 384–405, 1934.

E. E. Burger, *Gen. Elec. Rev.*, Vol. 37, pp. 93–99, 1934.

PART II

Principles of Television

CHAPTER 6

THE FUNDAMENTALS OF PICTURE TRANSMISSION

The problem of television, say ten or fifteen years ago, with the earlier mechanical systems, was to obtain an image which might be interpreted as a picture. The criterion of merit at that time was recognizability of the subject matter. Today, owing largely to the advent of cathode-ray television systems, a recognizable picture can be taken for granted, and the problem is one of obtaining a high-definition picture.

The term high-definition picture originated when it became apparent that it was not only desirable, but also possible, to obtain a picture having a much larger number of scanning lines and, consequently, higher resolution than heretofore thought feasible. Since then the term has come to mean not only a picture with a large number of scanning lines but also with low flicker level, correct contrast, sufficient brightness, and high "signal-to-noise ratio." In other words, a high-definition picture is a picture having a high degree of excellence.

A light image, produced, for example, by a *camera obscura*, is continuous over a two-dimensional surface and in time. The brightness is a function of the three coordinates, x, y, and t. It is fundamentally impossible to transmit such a picture over a single communication channel, that is, a channel such that the magnitude of the transmitted signal is a function of time alone. Since this limitation applies to both wire and radio communication, it presents the first major obstacle that must be overcome in the transmission of pictures. It is only through the sacrifice of something of the picture that the transmission of television images is possible.

The human eye suffers from certain limitations, and consequently the maximum picture "excellence" sensed by the observer is determined physiologically. The aim of television is to reproduce, if possible, a picture whose excellence is equal to this maximum excellence which can be observed.

The purpose of this chapter is, therefore, to consider what the requirements are to produce such a picture; to interpret these requirements in terms of television picture transmission; and, finally, if all the requirements cannot be met by a practical transmitted picture, to determine how best to dispose the resources available in order to obtain the most satisfactory picture.

75^2 picture elements 150^2 picture elements

225^2 picture elements 375^2 picture elements

original

Fig. 6.1.—Relation between Picture Perfection and Number of Picture Elements.

Advantage is taken of the finite resolving power of the eye and of the persistence of vision in order to make possible the transmission of pictures over a single channel.

Because of the limited resolving power of human sight, a given picture can convey only a finite amount of information. In other words, if a picture is considered as made up of small elements of area, each uniform in brightness, there is a limit beyond which any further decrease in size and increase in number of elements no longer improves the picture. This can be readily seen from Fig. 6.1. It is a fact that is made use of in all half-tone printing processes. Thus the reconstruction of any stationary picture at some remote receiver requires the transmission of a finite amount of information.

It is well known that, if pictures taken of an object in a consecutive series of positions are observed in sequence at a rate of more than twelve or sixteen pictures per second, the image gives the illusion of continuous motion. This is, of course, the familiar phenomenon utilized in producing motion pictures.

From the above it can be concluded that to reproduce a moving scene with all the detail of structure and motion that can be accepted by human vision requires the transmission of information at a finite rate only, and therefore can be accomplished over a single channel.

6.1. Basis of Television Transmission. The problem of the transmission of a picture of finite detail and discontinuous or limited motion has no unique solution. Of the several possible methods, that involving the process of scanning has been chosen as the most practical, at least for the present. This process consists of moving an exploring element or spot over the image to be transmitted in a periodically repeated path covering the image area. The exploring element is so constructed that it generates a signal which indicates the brightness (either instantaneous or averaged over a short time) of its instantaneous position. This signal is transmitted over the communication channel to the reproducing spot, whose brightness is controlled by the signal. The reproducing spot moves over the viewing screen in a path similar to, and synchronous with, that of the exploring element. Thus the reproducing spot reconstructs at the viewing screen, both in magnitude and position, the brightness distribution on the image area.

The path covering the image area and the viewing screen need not be continuous, but may be a series of straight parallel lines as shown in Fig. 6.2. This, in fact, is the most common form of scanning. There is, of course, a theoretical infinity of paths which might be chosen to cover the areas in question, for example, spiral scanning, sinusoidal scanning,

etc. In what follows, unless otherwise stated, straight-line scanning will be assumed.

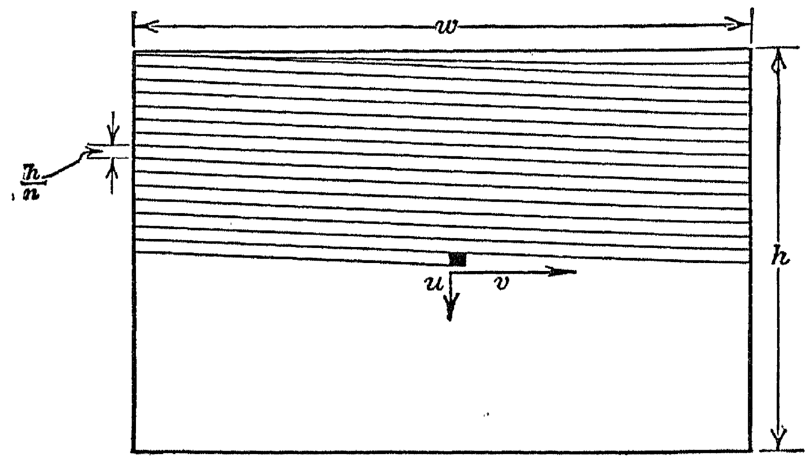


Fig. 6.2.—Scanning Pattern.

The complete system can be schematically represented by the five simple elements shown in Fig. 6.3. These elements are:

1. Scanning pattern on image field.
2. Scanning pattern on viewing screen.
3. Exploring element.
4. Reproducing spot.
5. Communication channel.

Roughly speaking, the geometrical correctness of the reproduced picture is dependent upon the scanning patterns; the tonal values, i.e., contrast,

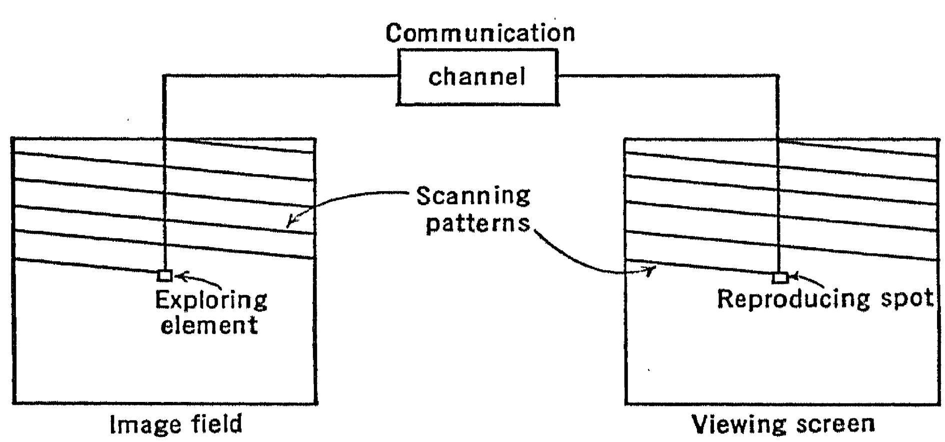


Fig. 6.3.—Functional Representation of a Television System.

etc., upon the communication channel; the range of brightness upon the reproducing spot; and the resolution upon all five of these factors. Actually this generalization is too broad to be of much value in considering a physical system of television, particularly because under it a portion

of the communication channel is included in both pickup and viewing means. However, for purposes of analyzing the transmission and reconstruction of a picture, this separation of functions is useful, and, once the analysis has been completed, the results can be fitted into the physical system.

In an actual cathode-ray system the exploring spot and image scanning pattern are in the pickup device, for example, the Iconoscope, where the exploring spot is formed by a cathode-ray beam deflected over the scanning pattern by a suitable electrostatic or magnetic arrangement. The reproducing spot and viewing screen are in the cathode-ray viewing tube, e.g., a Kinescope, where an electron beam also serves as the scanning element. The communication channel extends into the pickup device and includes the chain of video amplifiers, the radio transmitter, the medium through which transmission takes place, the receiver with its amplifiers, and, finally, the control element in the viewing tube.

6.2. General Considerations. The dependence of geometrical correspondence between image and reproduction upon the similarity and synchronism of the two scanning patterns requires no explaining. For all practical purposes, geometrical distortion can be due only to faults in the patterns. A slight distortion caused by faults in the communication channel, such as a variable time delay of the signal, is possible, but such faults are so much more serious from the standpoint of destruction of tonal values that geometric distortion from this cause may be ignored. In the present discussion it will be assumed that the scanning patterns at the receiving and transmitting end are identical; how this similarity in time and form is maintained, and the effects of dissimilarities, will be reserved for a later chapter.

In order to reproduce exactly the scene being televised, there should be a one-to-one correspondence between the brightness of the reproducing spot at any instant, and the brightness of the scene at the point corresponding to the position of the exploring element. In other words, the brightness of any point in the scene, and that of the corresponding point of the received picture, must bear the relationship shown in curve *a* of Fig. 6.4. This relation depends upon the sensitivity of the exploring element, the transmission characteristics of the electrical equipment, and the response of the reproducing spot, but will here be lumped as a property of the channel. Actually such a linear relationship may not be possible, or even desirable, in a practical television system. In many scenes which it might be of interest to transmit, the range of brightness is far beyond the capabilities of any of the existing reproducing systems. Furthermore, because of the difference in size of the scene and its picture, the lack of color and plasticity of the picture, and the conditions under

which it is viewed, a picture with brightness characteristics shown by curve *b* or *c* of Fig. 6.4 may yield a picture which is psychologically more accurate. In addition, the limited brightness range or contrast range may make curve *d* preferable. These curves hereafter will be referred to as the amplitude response curves of the system, and the slope of the curves as the gain of the channel. Photographic technique yields some clues to the answers of questions pertaining to the conditions giving best pictorial representation. This matter will again be considered in section 6.5.

It has been pointed out that, in order to maintain the illusion of continuous movement, the separate pictures showing the different steps of the motion must follow one another rapidly. Sixteen frames per second is sufficient for most pictures, although some types require a higher rate. The motion-picture industry has adopted 24 frames per second as a standard. This rate, possibly higher than absolutely necessary, insures smooth motion for all types of pictures. Instead of 24, 30 frames per second has been chosen as the most satisfactory repetition rate for television in this country, not because this high rate is essential for continuity of motion, but because it bears a simple relation to the frequency of commercial power. The reason why this relation is necessary will be taken up later. Actually 30 frames per second, while fulfilling all requirements of continuity of motion, is not a high enough frequency to avoid flicker effects. To meet this situation, interlaced scanning, a special form of straight-line scanning explained in section 6.4, is used, making unnecessary any greater frame frequency.

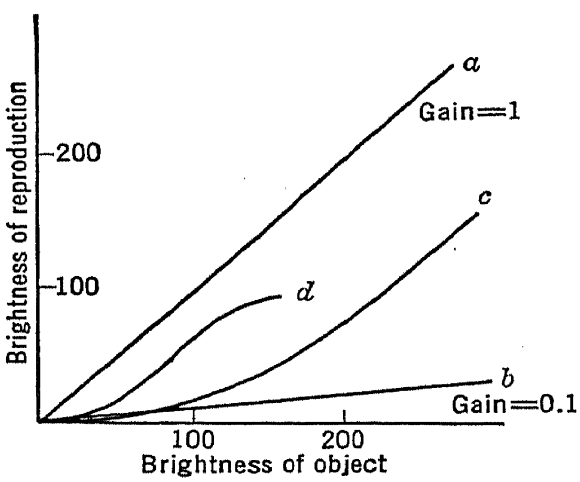


FIG. 6.4.—Relation between Brightness Variation in Object and Reproduction.

An analysis of resolution or definition is not simple, but is of sufficient importance to warrant considerable attention. Therefore, in addition to the following introductory remarks, the subject will be enlarged upon in sections 6.6, 6.7, and 6.8.

Vertical and horizontal resolution bear distinctly different relations to the mechanism of transmission. Vertical resolution depends only upon the number of lines making up the scanning pattern and upon the size of the scanning spots. The nature of the communication channel in practice plays no direct role in determining resolution vertically. Horizontal resolution, on the other hand, depends upon the channel and upon the

spot sizes and shapes. It is related to the number of lines only through the need of a particular rate of repetition of the pattern.

That the size of the exploring and reproducing spots affects the picture definition, and that small spots are required for both horizontal and vertical resolution, is self-evident from the mechanism that has been postulated for picture transmission. In the vertical direction it is obvious

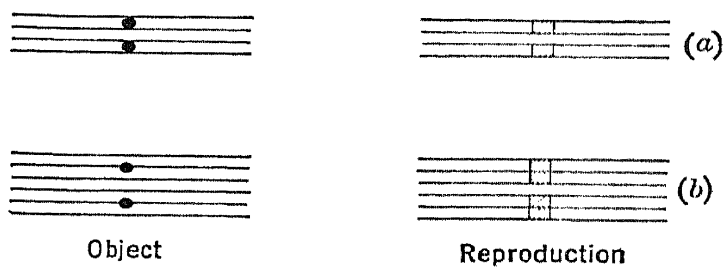


FIG. 6.5.—Vertical Resolution for Various Object Positions.

that two points cannot be resolved unless separated by two or three lines, as indicated in Fig. 6.5. To a first approximation, the resolution is equal to the separation between lines.

The nature of the communication channel affects horizontal resolution because of the fact that any physical channel is limited in the frequencies it can pass. In order to transmit shading in the vertical direction, the channel must pass the low frequencies; in fact, it will be shown that fre-

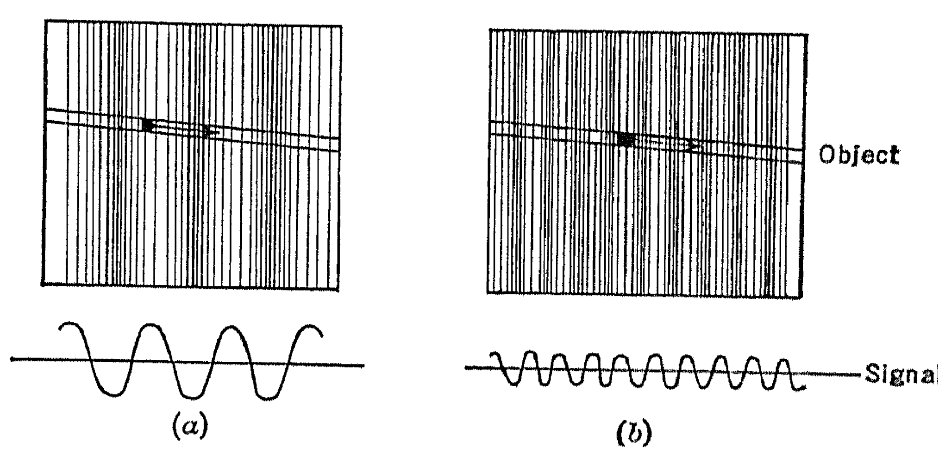


FIG. 6.6.—Frequencies Required to Transmit Vertical Line Patterns of Different Periodicities.

quencies down to, and even below, frame frequency must be transmitted for information about the background of an image whose average illumination is changing. The lower limit to the band passed, then, is approximately frame frequency, and therefore the bandwidth which must be passed by the channel is essentially equal to the frequency required to give the horizontal resolution needed. Qualitatively this frequency can be estimated by considering the signal generated by an image consisting of vertical lines as shown in Fig. 6.6. The scanning pattern will be as-

sumed to be made up of n lines, with a repetition rate of N frames per second. Then the time required by the beam to traverse one horizontal line is:

$$t = \frac{1}{n \cdot N} \text{ seconds.}$$

If the image consists of m vertical lines, the signal will go through m maxima and minima during each horizontal sweep, or, in other words, maxima will follow one another at a frequency of

$$f = \frac{m}{t} = m \cdot n \cdot N \text{ cycles per second.}$$

Obviously, as m increases this frequency becomes higher. If the image of the m lines is to be reproduced, the channel must pass all frequencies up to $m \cdot n \cdot N$. Conversely, if the channel will transmit frequencies up to $m \cdot n \cdot N$, black and white lines separated by a distance equal to $1/2m$ of the width of the field can be resolved. If the horizontal and vertical resolution is to be equal for a field whose width is w and whose height is h, the relation

$$\frac{h}{n} = \frac{w}{2m}$$

must be fulfilled, and the frequency band which must be passed by the channel is:

$$f = \frac{1}{2}\frac{w}{h}n^2N \text{ cycles per second.}$$

Actually the factor ½ does not yield the best results, as will be shown by a more detailed study.

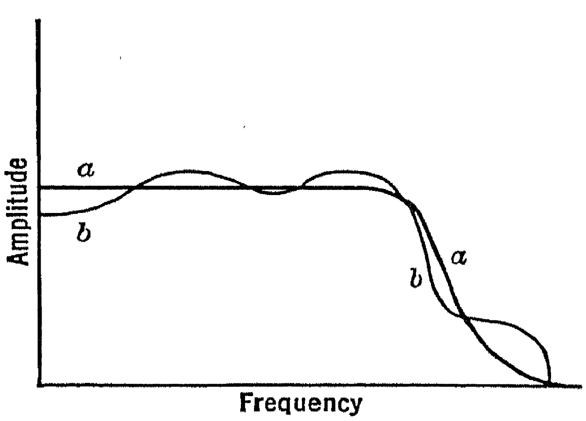


Fig. 6.7.—Amplitude as Function of Frequency for Two Different Channels.

The frequency requirement of the channel was derived from a very special case. Before the conclusions can be accepted as correct, an analysis, based on a Fourier expansion of a generalized image, must be made. However, on the basis of the above, a tentative conclusion that the gain of the channel should be independent of frequency from N cycles to a frequency of approximately $1/2(w/h)n^2N$ is not unreasonable. In other words, the frequency response characteristics should be those of curve a, Fig. 6.7, rather than those of curve b.

A final factor which must be considered in this survey is the length of time required by the signal to traverse the channel. In any physical system there will be a delay between the entrance of the signal into the channel and its emergence. If this delay is constant for all frequencies, it merely has the effect of displacing the reconstructed image on a scanning pattern which is exactly simultaneous with the transmitting pattern. Actually, since the scanning pattern is synchronized by a signal transmitted over the same channel that carries the picture signal, the pattern and picture are equally delayed so that this shift does not occur in reality. However, if the time delay is not constant, but varies with frequency, the high-frequency components of the picture will be displaced

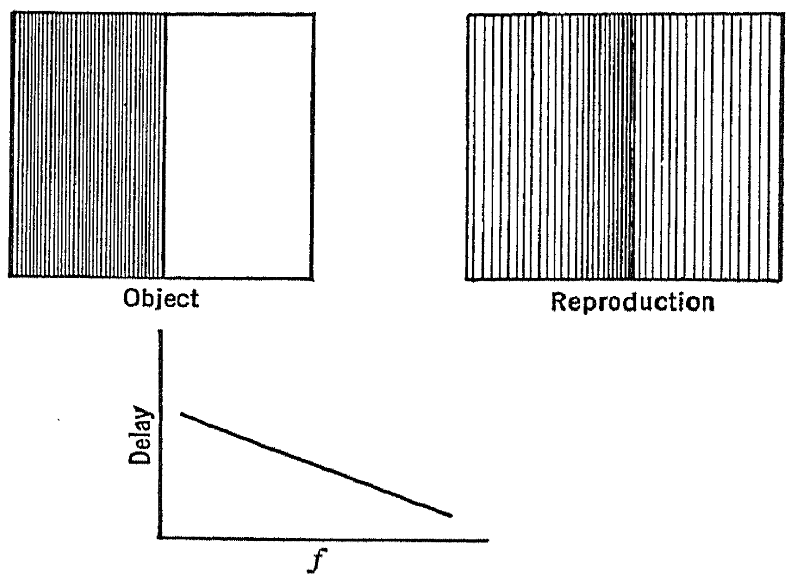


Fig. 6.8.—Effect of Non-Uniform Time Delay.

with respect to the low-frequency components, resulting in a serious distortion of the tonal values of the reproduction. This is illustrated in Fig. 6.8. Thus the communication channel not only must have a linear amplitude response and pass a wide frequency band at constant gain, but also must have a uniform time delay.

6.3. Brightness. From a qualitative point of view, it is immediately evident that the brightness of a picture and its excellence are related. Below a certain brightness, not only is a picture tiring to watch, but also it suffers loss of resolution. As the brightness is increased, eventually the detail becomes saturated; that is, all the detail which is above the visual resolving power of the eye can be seen. Even at this light level the picture will be tiring to watch for any considerable length of time, as anyone who has tried to read under conditions of too low illumination will be well aware. A further increase in light brings the brightness to a level such that the picture can be viewed with comfort. The range of

brightness over which the eye can accommodate itself with ease is very broad, but if the light level is increased too far, fatigue and discomfort again become apparent. This final region is of no concern to the television engineer at the present state of the art.

The relation between the ability of the eye to resolve and the surface brightness of the object is shown in Fig. 6.9. From this curve it is evident that the eye has reached its maximum acuity at about 0.1 candle per square foot. The brightness necessary for comfortable viewing has been determined by exhaustive measurements on the part of the motion-picture industry. The conclusions presented in a report before an S.M.P.E. Convention are as follows: In a moderately darkened room, 3.8 candles per square foot, or 11.8 foot-lamberts, are required for complete comfort; however, a surface having a brightness of 2.7 foot-lamberts or more can be viewed without serious discomfort for a period of several hours, and even as low a surface brightness as 1 foot-lambert can be watched with some degree of comfort in a well-darkened room.

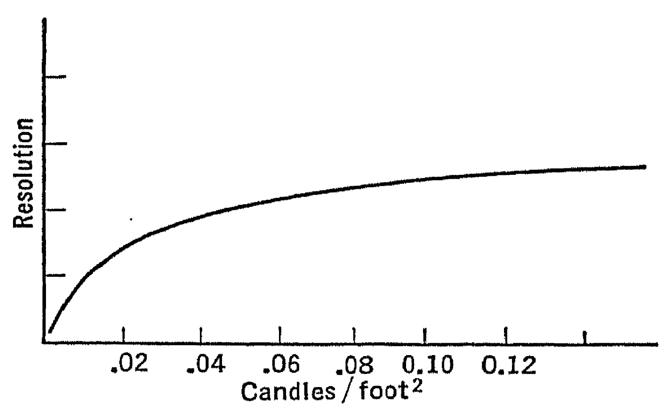


FIG. 6.9.—Visual Resolution as a Function of Brightness.

For a home television receiver which is designed to be used in a moderately lighted room, the screen brightness should be somewhat greater than the 12 foot-lamberts stated above. Practice shows that the 18 to 20 foot-lamberts, which can be readily obtained on the screen of a cathode-ray viewing tube, are ample.

6.4. Flicker. Even before the advent of motion pictures, the behavior of the visual sense towards an object which is periodically illuminated had been investigated. Since their advent the subject has become one of considerable practical importance, and consequently has been rather completely examined. The general conclusions are of interest to the television engineer, although the details, because of the rather special nature of the intermittency involved in television, are not particularly pertinent.

It was found that, although the eye is unaware of discontinuity of motion at frequencies above about 15 cycles per second, it nevertheless could detect flicker at very much higher frequency. The threshold frequency for awareness of flicker is a function of brightness of the object, color of the light, relative duration of light and dark, and the position on the retina where the light image falls. Fig. 6.10 illustrates the results

of a series of tests of flicker threshold as a function of brightness for different relative amounts of light and dark.* The length of the illuminated period is given in degrees, the entire period of light and dark equaling 360°. The light is averaged over the period, to correspond to the reaction of the eye to intermittent light. These tests were made with white light and with the image covering a large area of the retina. Other tests have shown that the threshold frequency is a maximum for light whose color is such that it produces the maximum visual sensation and decreases toward the red and blue, and, also, that the sensation of flicker at a high frequency is less at the edges of the retina than at the center.

The scanning process introduces a type of flicker which is sufficiently different from that described above that the results of these tests cannot

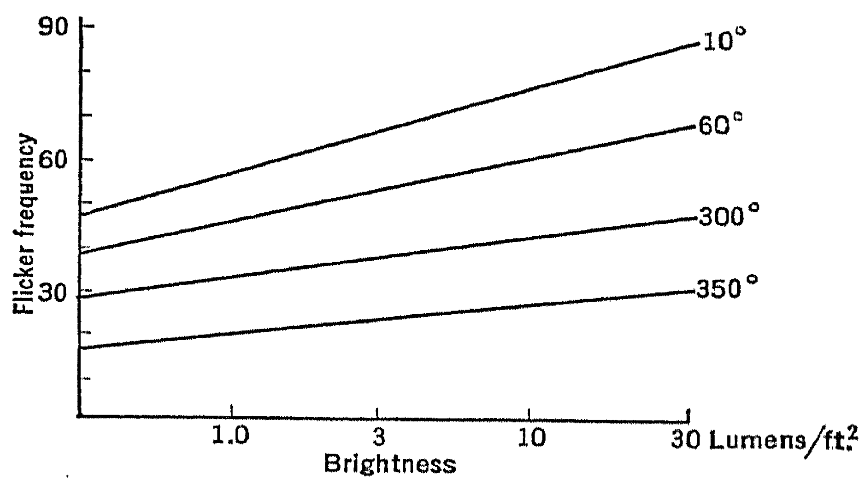


FIG. 6.10.—Relation of the Threshold Frequency of Flicker to Brightness.

be used, a priori, to formulate the requirements necessary to insure the elimination of flicker from the television viewing screen.

Reconstruction of the televised image on the viewing screen is accomplished by the motion of a bright modulated spot traversing the series of straight parallel lines making up the scanning pattern. Therefore, except for the very short time which may be required for the spot to return to its original vertical position (known as the return time), some point of the screen is always illuminated. However, if a given small area of the screen is considered, it is obviously illuminated intermittently for a very short interval at scanning frequency. For most practical cathode-ray television viewing tubes the duration of the interval for which the screen is illuminated is a function of the decay characteristics of the phosphor used on the screen, rather than the actual length of time the scanning spot is on the area under consideration. The most commonly used phosphors have a decay time which is long enough so that a given area re-

* See Engstrom, references 3 and 4.

mains luminous for a time which is long compared to that taken by the spot to traverse a horizontal line. Thus, as far as flicker is concerned, the process of picture reconstruction may be considered as being the result of a bright horizontal strip moving vertically across the screen at frame frequency. The leading edge of this strip has the maximum brightness, the luminosity decaying gradually towards the trailing edge. In order to determine the reaction of the visual senses to this type of flicker, tests were made using the projected image of continuously moved film whose frames had the density distribution shown in Fig. 6.11. The results of these are shown in Fig. 6.12. The general qualitative resemblance of these results with those of Fig. 6.10 is evident. From the information given in these tests it can be concluded that the scanning field frequency must be about 50 cycles per second or higher to avoid all sensation of flicker on a bright picture.

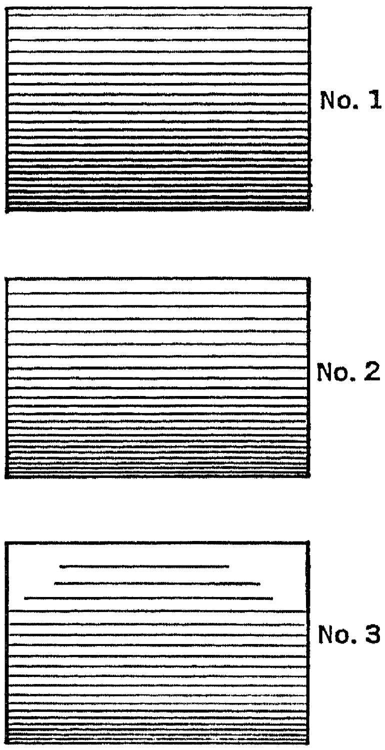


FIG. 6.11.—Test Films Used to Determine Threshold of Flicker for a Scanning Pattern.

This frequency is much higher than is required for continuity of motion, and therefore requires an unnecessarily large frequency band of the communication channel. Economy of bandwidth can be retained without reducing the effective field frequency by means of interlaced scanning.

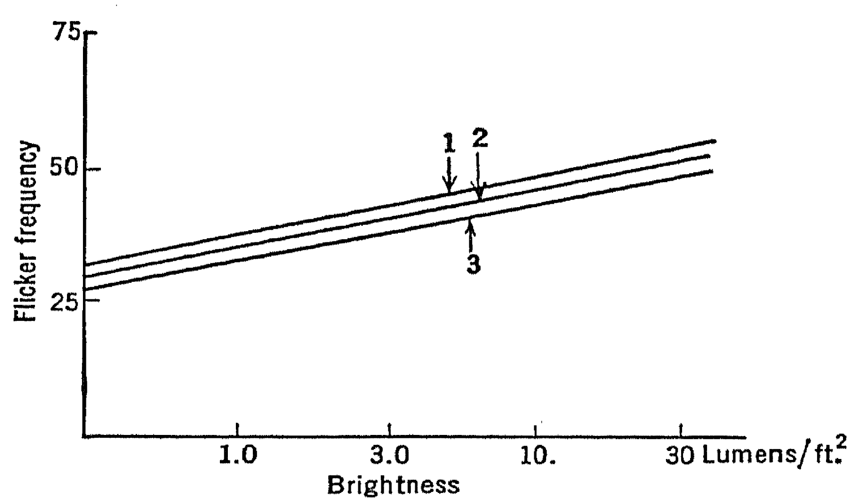


FIG. 6.12.—The Threshold Frequency of Flicker for a Scanning Pattern.

Interlaced scanning is a form of straight-line scanning. The spot, instead of moving across the horizontal lines in the sequence 1, 2, 3, 4, · · ·, covers the odd-numbered lines first, then the even-numbered

lines, i.e., 1, 3, 5, · · ·; 2, 4, 6, · · ·. This process of scanning is illustrated in Fig. 6.13. It will be seen that the scanning beam has to complete two scanning patterns, displaced by the width of one line, in order to cover the entire field. Therefore the field frequency is twice the frame frequency.

This form of scanning is not restricted to covering every second line in the first vertical sweep and the intervening lines on the second sweep.

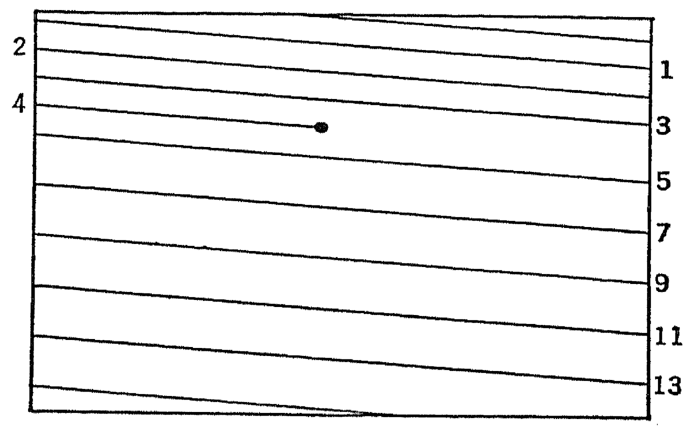


Fig. 6.13.—Interlaced Scanning.

The scanning may be interlaced in such a way that the first sweep covers every third or every fourth line, etc.; that is, the sequence may be:

1, 4, 7, 10, · · ·; 2, 5, 8, · · ·; 3, 6, 9, · · ·

or:

1, 5, 9, 13, · · ·; 2, 6, 10, · · ·; 3, 7, 11, · · ·; 4, 8, 12, · · ·

or, in general:

1, $k + 1$, $2k + 1$, · · ·; 2, $k + 2$, $2k + 2$, · · ·; 3, $k + 3$, $2k + 3$, · · ·; · · ·; k, $2k$, · · ·.

Interlaced scanning, as applied in most practical systems, employs a frequency which is twice the repetition frequency of the picture. Tests show that at optimum viewing distance, namely, a distance such that the line structure is just below the limit of resolution, the field frequency of interlaced scanning and frame frequency of regular scanning have identical relationship to the perception of flicker. Therefore, a field frequency above 50 cycles with interlaced scanning gives a flicker-free reproduction. However, if the individual lines can be resolved, a slight interline flicker or weave may be perceptible in a bright picture. In the United States, for reasons connected with the 60-cycle standard of commercial power, a repetition rate of 30 frames per second and a field frequency of 60 cycles per second have been adopted, while in England and in conti-

nental Europe, where 50-cycle power is standard, frame and field frequencies are 25 and 50 cycles, respectively.

6.5. Contrast. The relation between brightness of corresponding points of the original image and the reproduced picture is important in determining the merit of the picture. This relationship is not, however, as critical as the geometry and resolution of the reproduction, and because of this it is a phase of picture transmission which has not until recently received much attention.

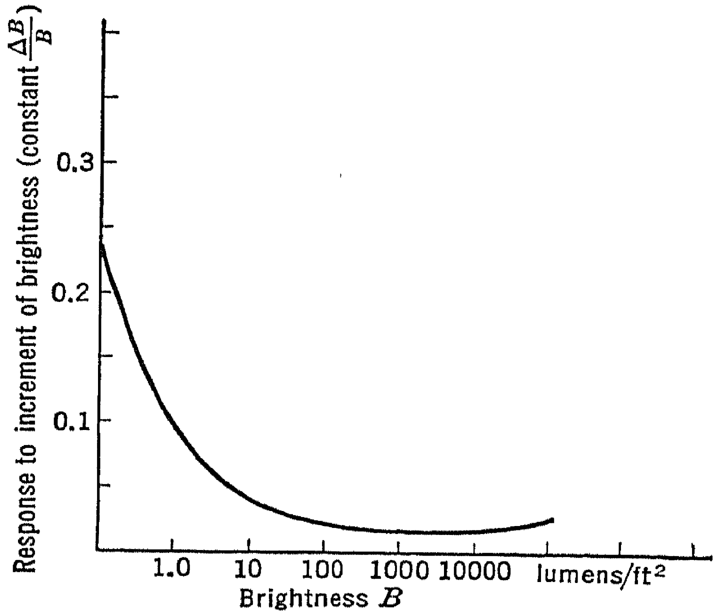


FIG. 6.14.—Visual Response to Incremental Increases in Brightness.

As has already been pointed out, the brightness of the image point for a given object illumination depends upon the amplitude response of the communication channel, which includes the sensitivity of the exploring element and the characteristics of the reproducing spot. By properly regulating these factors, any desired brightness relationship over a wide range can be obtained.

If an object consisting of adjoining strips illuminated in such a way that the brightness of successive strips is related by

$$\frac{B_{n+1} - B_n}{B_n} = \frac{\Delta B}{B} = \text{const.}$$

is observed, the increments of visual sensation which are produced, when plotted as a smooth curve, will be as shown in Fig. 6.14. This relation between the sensation of increase in brightness and brightness is known as Fechner's law. For the range of brightness over which objects can be observed without fatigue, that is, from about 0.3 candle per square foot

to 100,000 candles per square foot, the sensation received for an increment of relative brightness $\Delta B/B$ is practically independent of brightness.

An alternative statement expressing this same observation is that visual response is proportional to the logarithm of the brightness, a relationship shown in Fig. 6.15. This property of the response has been confirmed by many different types of tests. Because a brightness increment ΔB with equal $\Delta B/B$ produces an equal visual response over a wide range of brightness, it is very useful as a means of comparing the relative brightness of various parts of an image, and as such has been given the name visual contrast.

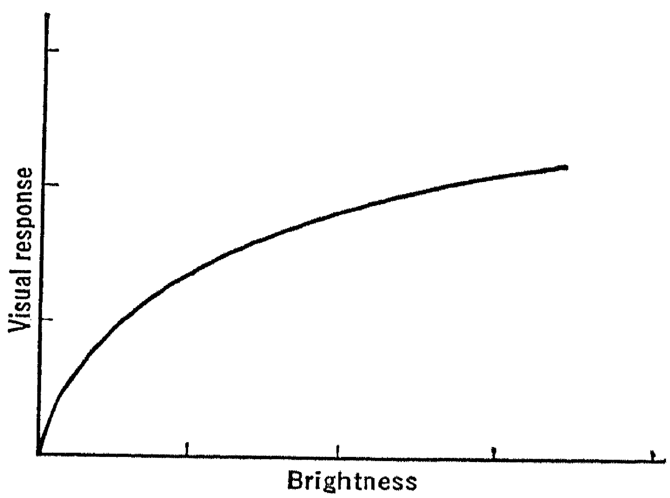


FIG. 6.15.—Visual Response as Function of Brightness.

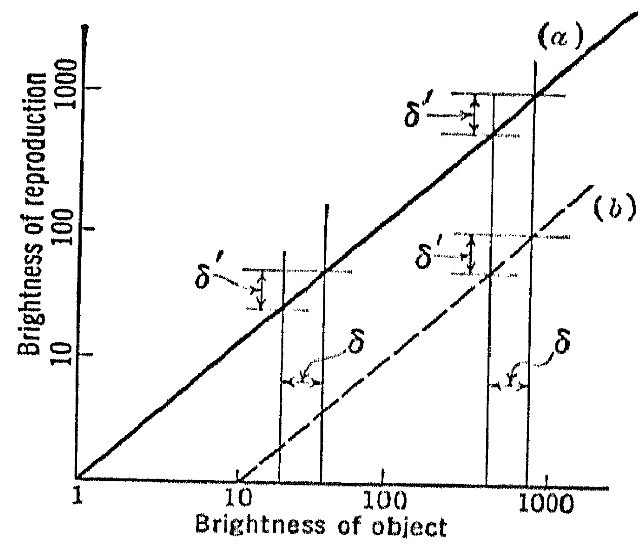


FIG. 6.16.—Brightness and Contrast.

In comparing the contrast properties of two images, for example, that of a transmitted and received image, it is useful to construct a logarithmic plot of the brightness of corresponding points. When the response of the channel is linear and the gain unity, the brightness of transmitted and received image points plotted in this way lies on a straight line at 45° to the abscissa, as shown in curve (*a*) of Fig. 6.16. A change in gain does not alter the slope of the line, but merely moves it up or down, as is the case with curve (*b*) representing a gain of 1/10. A given contrast, irrespective of the average brightness, between two areas is represented by a constant interval. For example, the interval δ in the figure represents a contrast of 0.2 in the image. Since both the lines (*a*) and (*b*) make an angle of 45°, a given δ of the transmitted image results in an equal interval δ' along the ordinate representing the received image brightness. Therefore the contrast in the two images will be equal.

When the curve representing the received and transmitted brightness is a straight line, making an angle either more or less than 45° with the

abscissa, as shown by curves (*b*) and (*c*), Fig. 6.17, the contrast in the two pictures will no longer be the same. However, the relative contrast of various parts of the transmitted image will be the same as for those in the received image. That is, if the contrast between two objects in one portion of the transmitted image is twice that of two objects in some other portions, the contrast between the first pair of objects will be twice that of the second in the received image. This relationship is illustrated diagrammatically in the figure. If the line representing the response makes an angle greater than 45° the contrast in the received image is greater than that in the transmitted image, and the received image ap-

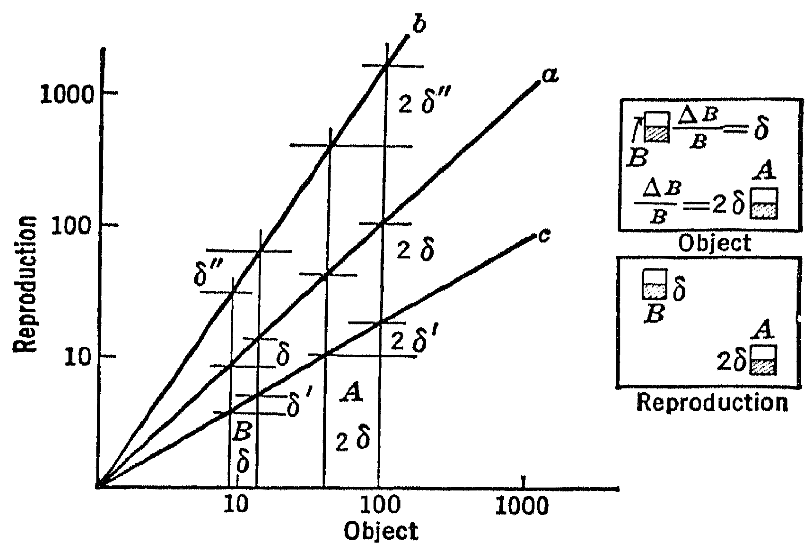


FIG. 6.17.—Relative Contrasts in Object and Reproduction.

pears crisp or harsh, whereas if the angle is less than 45° the rendition is soft or flat.

In photographic practice the slope of the response line has been given the name gamma (γ). Thus for identical contrasts γ is unity; if it is greater or less in the reproduction, γ is greater or less than one. The concept of γ is very useful in photography because in the process of development the linearity of the logarithmic plot between the reproduction transparency and the brightness of the original is preserved, but the slope or γ of the curve changes with the various factors, such as time and temperature, influencing development. This same nomenclature may eventually prove useful in the field of television, although in the physical communication channel there is no element which automatically preserves the linearity of logarithmic response, nor is it at present known whether, if it proves desirable to use a non-linear channel response, the output should be a fixed power of the input as implied by a constant gamma.

There are two controls at the ordinary cathode-ray receiver which in-

fluence the brightness and brightness relationships in the reproduced picture. These are the amplifier gain control and the background control. If the brightness on the viewing screen is proportional to the signal voltage, and the rest of the system has a linear response, these two controls have the following effect. The former merely operates to regulate the overall gain of the channel, and as such does not influence the contrast relations in the picture. The latter adds a constant amount of light to every part of the picture. These two effects are illustrated in Fig. 6.18. To show the effect on the contrast of the reproduced picture, these curves are replotted on a logarithmic scale. It will be seen that by the combined use of the gain and background controls the contrast of the reproduction can to some extent be regulated. This is illustrated in Fig. 6.19.

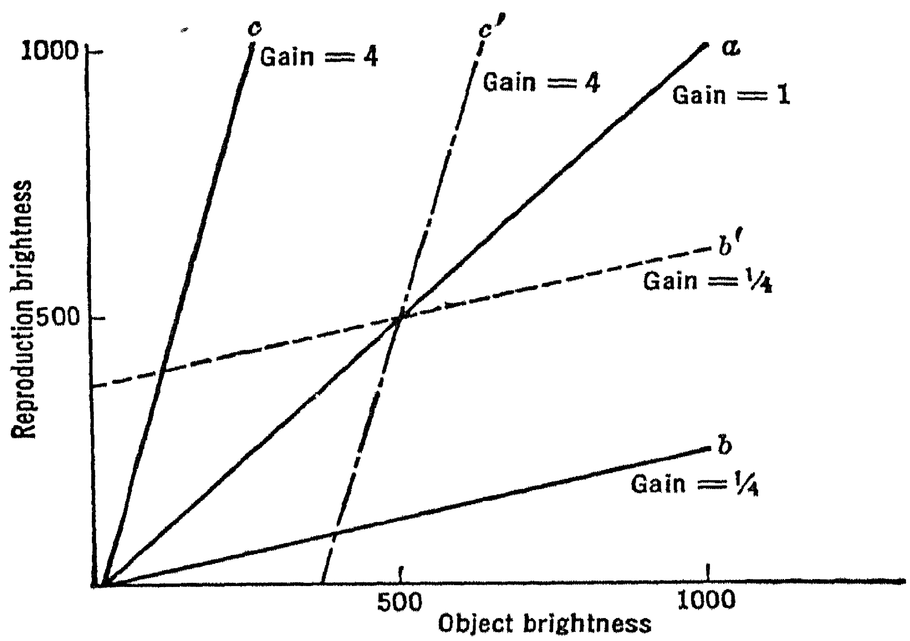


FIG. 6.18.—Effect of Gain and Brightness Controls.

On the other hand, if the cathode-ray reproducing tube has an exponential response, and the signal picked up by the exploring spot is compressed so that its amplitude is proportional to the logarithm of the brightness at every point, then the channel gain controls the γ, i.e., the contrast, of the reproduction, while the "background" control actually regulates the effective gain of the system.

The present practical systems are a compromise between the two types described above.

The range of contrast which is possible in the reproduced picture is usually expressed as the ratio of the maximum to the minimum brightness of the screen, and is often much smaller than the range of contrast in the image being transmitted. This limited range is due to the fact that the minimum brightness is determined by various types of optical and

electron scattering in the system, and cannot be made arbitrarily small, while the maximum brightness is limited by fluorescent screen materials, electron guns, apertures of mechanical scanning systems, surface brightness of light sources, etc. Cathode-ray viewing tubes normally have a contrast range of 50 to 100. The smallest steps in contrast $\Delta B/B$ are determined by the random fluctuations or noise in the electrical systems involved in the pickup, channel, and receiver. Both range of contrast and limitations imposed by noise will be considered again in other chapters.

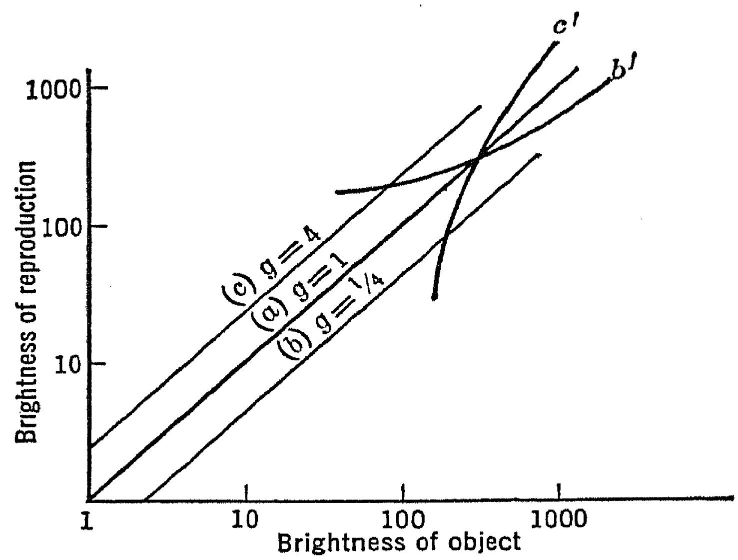


FIG. 6.19.—Effect of Gain and Brightness Control on Relative Contrast.

6.6. Resolution. Occasionally a picture which is essentially continuous is encountered, but most of those viewed are grainy in nature. Pictures of the latter type must be divided into two classes: those in which the grain is regularly disposed, and those in which the individual grains are scattered at random. Another classification divides them into those in which the individual grains may have any shading from light to dark, and those in which only black grains and white grains are used and picture contrast is obtained by the relative density of the two kinds. These two classifications are mutually inclusive.

The scanned picture cannot be brought into the fold of either grainy or continuous, because along any line it is essentially continuous, while at right angles to this it is grainy in nature with a regular distribution of elements.

The question of resolution is relatively simple in a picture of regular grain, but becomes more complex for a continuous picture. Considering the former, it is evident that the grain, or picture elements, should be sufficiently small that the eye does not resolve the individual units. Numerous tests show that the average eye cannot resolve two points separated by a distance subtending less than 1 minute at the eye. Exceptional eye-

sight under ideal illumination can resolve as little as ½ minute, while eyes which cannot resolve less than 2 minutes of angle are fairly common. Actually this cannot be interpreted to mean that, where a regular pattern of picture elements is used, elements spaced at 1 minute can reproduce a picture whose limiting resolution is 1 minute of angle, because the components of the picture are randomly distributed. However, a regular pattern of elements with spacings which subtend an angle of about 1 minute of angle is capable of forming images which reach the limit of visual acuity for all but very special patterns (i.e., patterns having regularities which "beat" with the element regularity).

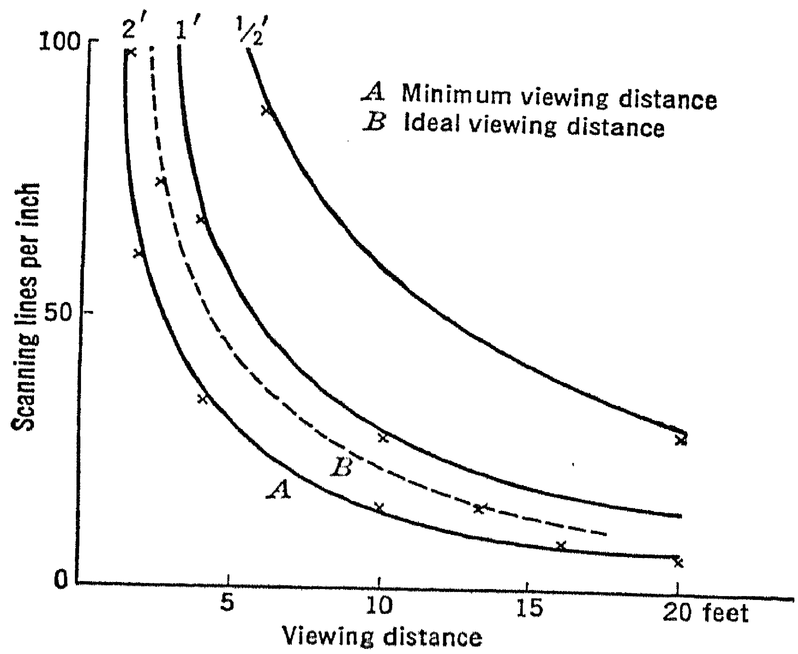


FIG. 6.20.—Relation between Picture Structure and Viewing Distance.

Measurements made by E. W. Engstrom,* using projected pictures whose structure simulates that of a television picture, give valuable information relative to visual acuity, image structure, and picture size. His conclusions are based upon the results of a number of tests by several observers on pictures having structure and structureless images. The findings from some of these tests, in terms of scanning lines per inch and viewing distance, are shown in Fig. 6.20. The general conclusion from these data is that the optimum viewing distance is such that the structure subtends an angle of about 1½ minutes. At this distance the structure becomes unobtrusive or invisible, and the detail appears to be the same as that of a structureless reproduction of a similar image viewed from the same distance.

From moving-picture practice and experience it can be concluded that a picture whose height subtends an angle of about 10 to 20 degrees at the

* See Engstrom, reference 3.

eye is a satisfactory compromise between being so small that it is fatiguing because of the effort which must be made to keep the attention fixed upon it, or so large that there is fatigue due to the motion of the eyes in following the action.

Taken together, these two factors, i.e., a total angle $\theta = 10°$ and a structure $\delta = 1.5'$, determine the minimum number of lines satisfactory for the scanning pattern. The number of lines should be approximately:

$$n = \frac{\theta}{\delta} = 400 \text{ lines.}$$

Actually 441 lines has been adopted for the present by the Radio Manufacturers Association as a standard for this country.

It has been stated that an equal horizontal and vertical resolution is to be desired, at least to the best of present knowledge. Since in the ver-

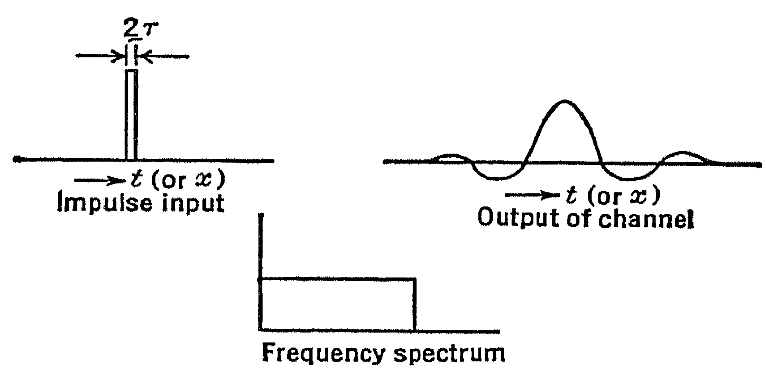


FIG. 6.21.—Reproduction of a Bright Bar by a Sharply Cutoff System.

tical direction the transition from black to white occurs in a distance equal at most to the separation between the centers of two lines, that is, in a distance $l = h/n$, this same separation should be attainable in the horizontal direction. If the image of a small illuminated point is transmitted, the resultant reconstructed image will be a bright spot somewhat larger than the original point. Its width, in order to meet the resolution requirements for the horizontal width, should not be more than $2l$. In order to determine the channel width needed to fulfill this requirement, it is necessary to make use of a Fourier integral. (Note: A Fourier series is actually required if the spot is stationary owing to the periodicity of the signal, but is more cumbersome to apply and lends nothing to the accuracy of the present calculation.)

The scanning beam in passing over the illuminated part gives rise to an impulse of very short duration which can be represented as shown in Fig. 6.21. The frequency spectrum is given by:

$$F(f) = \frac{2A}{\tau}\int_0^{\tau} \cos 2\pi f t \cdot dt = 2A\frac{\sin 2\pi f\tau}{2\pi f\tau} \qquad (6.1)$$

where 2τ is the duration of the impulse. If τ is small this represents an essentially constant band out to a very high frequency, the frequency limit increasing as the duration of the pulse decreases. The communication channel, however, transmits only a band of frequencies extending from approximately zero to a cutoff frequency f_c. Therefore the frequency spectrum of the pulse at the receiving end of the channel consists of a band of constant amplitude A extending from zero to f_c. The integral representing the resultant pulse is:

$$P(t) = 2A\int_0^{f_c} \cos 2\pi ftdf = \frac{A}{\pi}\frac{\sin 2\pi f_c t}{t}. \tag{6.2}$$

Fig. 6.21 illustrates the form of the pulse. The length of time between the point of maximum amplitude and the first minimum can be found by determining the derivative of $P(t)$ with respect to t and equating it to zero.

$$\frac{dP(t)}{dt} = 2Af_c\frac{\cos 2\pi f_c t}{t} - \frac{A}{\pi}\frac{\sin 2\pi f_c t}{t^2} = 0, \tag{6.3}$$

$$2\pi f_c t_0 = \tan 2\pi f_c t_0, \tag{6.3a}$$

$$f_c = \frac{4.4934}{2\pi}\frac{1}{t_0}.$$

The half width l of the reproduced image is equal to the product of the velocity of the reproducing spot and the time t_0. Therefore

$$t_0 v = l = \frac{h}{n};$$

and, since

$$v = Nnw,$$

it follows that

$$t_0 = \frac{h}{w}\frac{1}{n^2N},$$

and from this the cutoff frequency of the channel will be

$$f_c = \frac{4.493}{2\pi}\frac{w}{h}n^2N = 0.715\frac{w}{h}n^2N. \tag{6.4}$$

In this expression w/h is the aspect ratio of the picture, which may be taken as 4/3.

Equation 6.4, when evaluated for a 441-line picture having a repetition rate of 30 frames per second, gives a cutoff frequency of about 5½ megacycles.

Actually the assumptions made in the evaluation were much too simple to lead to an accurate estimate of the bandwidth. Two very important factors that were not considered at all are the properties of a physical spot, and the shape of the transmission band as it exists in reality. Before these factors can be added to the analysis of the bandwidth requirements the theory of scanning must be examined more deeply.

6.7. Theory of Scanning. The analysis of a picture into an electrical signal and the synthesis of this signal back into an image by the scanning are quite complicated. A very complete treatment of this subject was given by Mertz and Gray,* making use of harmonic analysis. The following discussion is based upon their treatment, but is necessarily seriously abridged because of lack of space, and therefore an interested reader will do well to refer to the original article.

The scanning of a stationary picture is a process of repetition; consequently the signal produced can be represented by the Fourier series

$$E(t) = \sum_{k=0}^{\infty} A_k \cos\left(\frac{2k\pi t}{T} + \theta_k\right), \tag{6.5}$$

where k is the order of the term, A_k the amplitude of the kth Fourier component, and T the frame period. It will be noted that the number and frequency of the Fourier components do not depend upon the light distribution in the picture, but only upon the scanning process. The coefficients A_k are, however, functions of the light distribution in the image being scanned.

The variable t can be expressed in terms of the position of the exploring spot with the aid of the relation:

$$\frac{t}{T} = \frac{x}{wn}.$$

The series can therefore be written in the form:

$$E(x) = \sum_{k=0}^{\infty} A_k \cos\left(2\pi\frac{kx}{wn} + \theta_k\right). \tag{6.6}$$

This can best be visualized by considering that the scanning process consists of moving the exploring spot over an array of identical images, as shown in Fig. 6.22.

The one-dimensional series just given, which is required by the channel, does not provide a means of determining whether this type of series can contain components of such a form that the picture can be reconstructed.

* See Mertz and Gray, reference 5.

In order to represent completely a two-dimensional image, it is necessary to use a double series. If $B(x,y)$ is the brightness at any point x, y, the image can be described by the following series:

$$B(x,y) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} A_{kl} \cos\left[2\pi\left(\frac{kx}{w} + \frac{ly}{h}\right) + \theta_{kl}\right]. \tag{6.7}$$

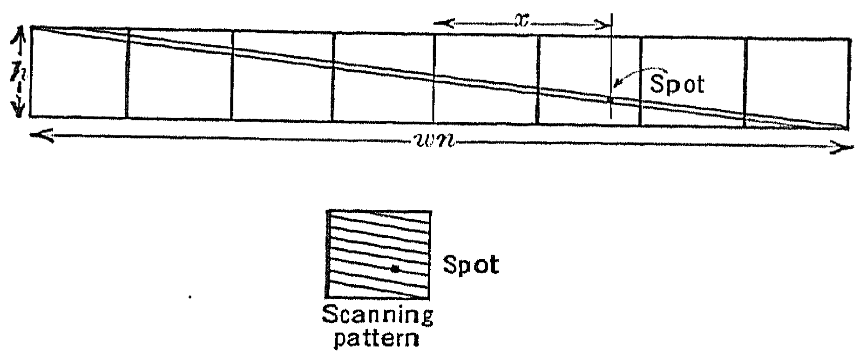


FIG. 6.22.—One-Dimensional Representation of Scanning.

This series represents an array of sinusoidal variations over the field, of all frequencies and in all directions. Examples of some of the components are shown in Fig. 6.23. Since in the scanning process both x and y are

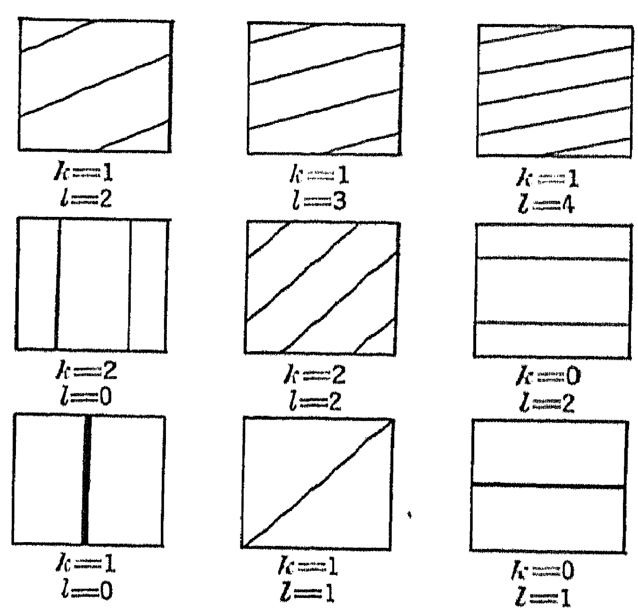


FIG. 6.23.—Components of the Fourier Series Used in Picture Analysis.

functions of time alone, the series given in Eq. 6.7 can be written in terms of this variable

$$E(t) = \sum_{k}\sum_{l} A_{kl} \cos\left[2\pi\left(\frac{kv}{w} + \frac{lu}{h}\right)t + \theta_{kl}\right], \tag{6.8}$$

where $vt = x$ and $ut = y$.

However, there exists, from the nature of the scanning pattern, a relation between the horizontal velocity v and the vertical velocity u, namely:

$$u = \frac{h}{w}\frac{v}{n},$$

and, furthermore,

$$v = wnN.$$

Therefore the frequencies f which appear as argument for the cosine terms are:

$$f = \frac{v}{w}\left(k + \frac{l}{n}\right) \tag{6.9}$$

$$= nN\left(k + \frac{l}{n}\right). \tag{6.9a}$$

Several interesting points may be noticed in this relation. When k is zero it represents a family of frequencies which are multiples of the frame frequency. All frequencies for which l is zero are multiples of the line frequency nN. The resulting spectrum may be thought of as groups of frequencies about the various multiples of line frequency, the members of each group being separated by multiples of frame frequency. Another feature of this relation, which will be considered at greater length presently, is that f can have the same value for several values of k,l.

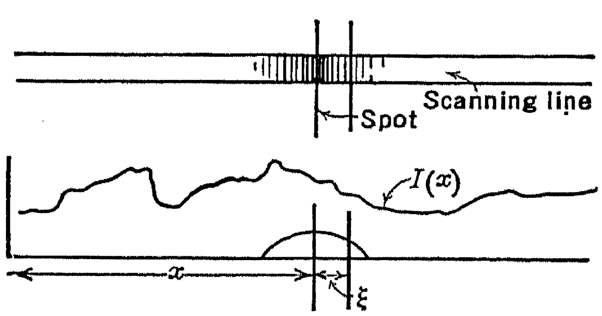


FIG. 6.24.—Scanning with a Finite Aperture.

So far both the exploring element and the reproducing spot have been considered as points. Actually, of course, these spots have a finite size which is comparable with the resolution of the device. The effect of the size is a filtering action on the higher frequencies. Returning again to the one-dimensional series given in Eq. 6.6, the quantitative effect of a finite exploring aperture can readily be shown. Fig. 6.24 represents the line being scanned, together with the finite aperture whose response, which varies over its length, is shown by the small rounded curve.

The picture intensity I along the scanned line is a function of the distance along the line. However, the amount of signal picked up from any point depends not only on the intensity I, but also upon the particular

region of the spot that lies on that point. Let x be the coordinate locating the center of the exploring spot, and ξ that locating a position within the spot itself, relative to the center of the spot. For simplicity it will be assumed that the exploring element is symmetric about the center and that its response is $S(\xi)$.

The intensity of illumination at point $(x + \xi)$ is:*

$$I(x + \xi) = \sum_{k=-\infty}^{\infty} A_k \exp\left[\frac{2\pi ik(x + \xi)}{wn}\right], \tag{6.10}$$

while the signal from the total area of the exploring spot at point x is:

$$E(x) = \int S(\xi)I(x + \xi)d\xi. \tag{6.11}$$

Therefore, substituting Eq. 6.10 in Eq. 6.11, the signal becomes:

$$\begin{aligned} E(x) &= \sum_k \int S(\xi)A_k \exp\left(\frac{2\pi ik(x + \xi)}{wn}\right)d\xi \\ &= \sum_k \int S(\xi) \exp\frac{2\pi ik\xi}{wn} A_k \exp\frac{2\pi ikx}{wn} d\xi, \end{aligned} \tag{6.12}$$

or

$$E(x) = \sum_k A_k Y(k) \exp\frac{2\pi ikx}{wn}, \tag{6.12a}$$

where

$$Y(k) = \int S(\xi) \exp\left(\frac{2\pi ik\xi}{wn}\right)d\xi = \int S(\xi) \cos\left(\frac{2\pi k\xi}{wn}\right)d\xi. \tag{6.13}$$

* While the cosine series is convenient for physical interpretation, some mathematical simplification is effected by the use of the exponential form of the Fourier series. Therefore the latter will be used, returning to the former only where practical interpretation makes it necessary. The relation between the forms can be seen from

$$2a \cos(x + \phi) = (ae^{i\phi})e^{ix} + (ae^{-i\phi})e^{-ix}.$$

Therefore

$$\sum_{k=0}^{\infty} a_k \cos\left(\frac{2\pi k(x + \zeta)}{wn} + \phi_k\right)$$

becomes

$$\sum_{k=-\infty}^{k=\infty} A_k \exp 2\pi i\left[\frac{k(x + \zeta)}{wn}\right],$$

where

$$A_k = (1/2)a_k e^{+i\phi_k};\ A_{-k} = (1/2)a_k e^{-i\phi_k}.$$

The factor $Y(k)$, called the aperture admittance, appears in the series in the same form as electrical admittance, since it is a function of frequency only in the one-dimensional case. The function is illustrated in

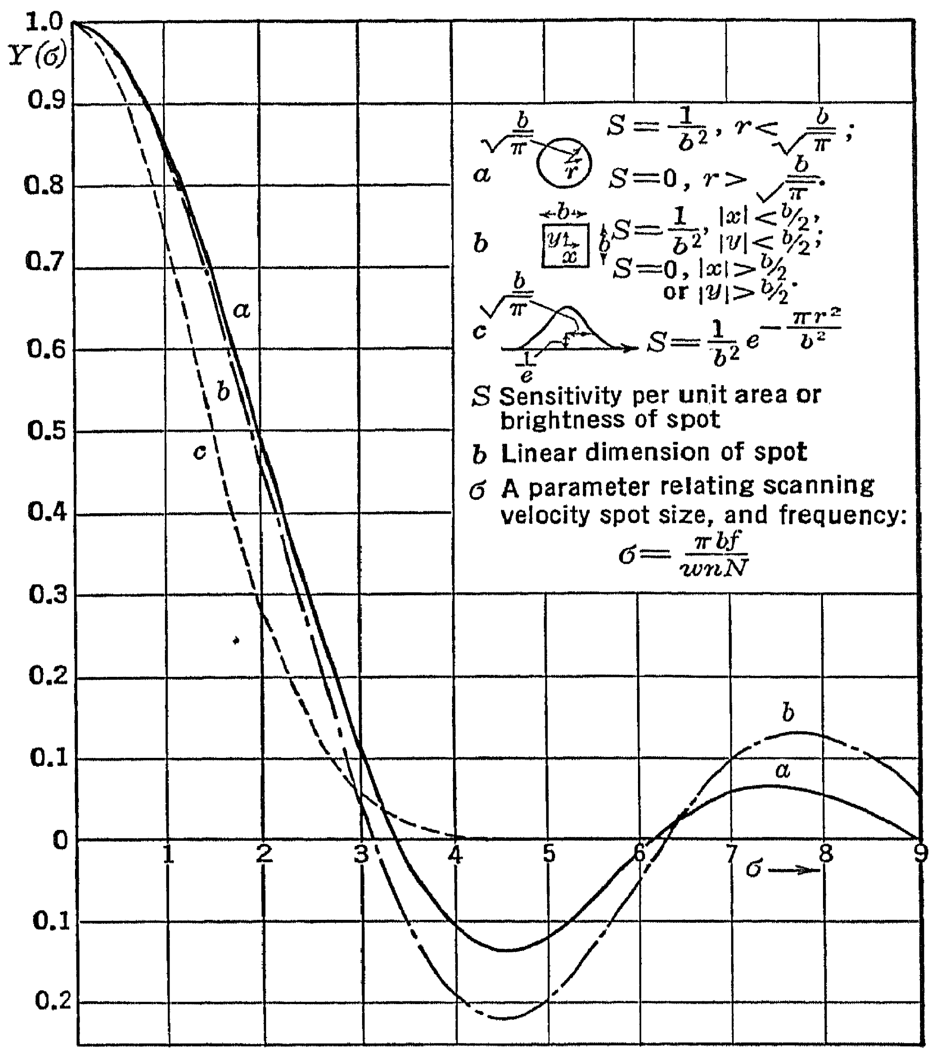


FIG. 6.25.—Aperture Admittance of Various Scanning Spots.

Fig. 6.25 for a uniformly sensitive round spot (a), square spot (b), and (c), one in which the sensitivity is given by $S = e^{-\frac{\pi r^2}{b^2}}$.

For the more complete two-dimensional analysis resulting in Eq. 6.8, the aperture admittance becomes:

$$Y(k,l) = \int\int S(\xi,\eta) \exp 2\pi i\left(\frac{k\xi}{w} + \frac{l\eta}{h}\right) d\xi d\eta \tag{6.14}$$

and

$$E(x,y) = \sum_k \sum_l Y(k,l) A_{kl} \exp 2\pi i\left(\frac{kx}{w} + \frac{ly}{h}\right) \tag{6.15}$$

Thus as a function of time the signal is

$$E(t) = \sum_k \sum_l Y(k,l) A_{kl} \exp 2\pi i n N\left(k + \frac{l}{n}\right)t. \qquad (6.16)$$

If the sensitivity of the exploring spot has symmetry about both ξ and η axes, $Y(k,l)$ becomes:

$$Y(k,l) = \int\int S(\xi,\eta) \cos 2\pi\left(\frac{k\xi}{w} + \frac{l\eta}{h}\right) d\xi d\eta. \qquad (6.17)$$

A very commonly occurring case is that in which the sensitivity can be broken into two factors which are each dependent on one variable only,

$$S(\xi,\eta) = S(\xi) \cdot S(\eta),$$

in which case the admittance becomes:

$$Y(k,l) = Y(k) \cdot Y(l), \qquad (6.18)$$

where

$$Y(j) = \int S(\delta) \cos 2\pi\left(\frac{j}{p}\right) \delta d\delta. \qquad (6.19)$$

The signal arriving at the receiving end of the system, neglecting for the moment the frequency response of the channel, is, of course, $E(t)$. This signal modulates the brightness of the reproducing spot which scans the viewing screen. The reproducing spot is finite in size, and not necessarily of uniform brightness over its area. Let $S'(\xi,\eta)$ be the function representing its brightness with respect to a point fixed on the spot, the primed symbols being used to designate properties of the receiving system.

The instantaneous brightness of a point on the screen will be:

$$S'(\xi,\eta)E(t) = S'(x - vt,\ y - ut)E(t). \qquad (6.20)$$

Let $B'(x,y)$ represent the brightness at a point x,y in the reproduced image. By the same reasoning as before, this may be expressed by a double Fourier series, as follows:

$$B'(x,y) = \sum_{k'} \sum_{l'} A'_{k'l'} \exp 2\pi i\left(\frac{k'x}{w} + \frac{l'y}{h}\right). \qquad (6.21)$$

To compare the received picture with that transmitted, the Fourier

terms of the two pictures may be compared. The coefficients $A'_{k'l'}$ of the image on the reproducer are:

$$A'_{k'l'} = \frac{1}{wh}\int_{-\frac{w}{2}}^{+\frac{w}{2}}\int_{-\frac{h}{2}}^{+\frac{h}{2}} B'(x,y) \exp - 2\pi i\left(\frac{k'x}{w} + \frac{l'y}{h}\right)dydx. \quad (6.22)$$

Using Eq. 6.20 to express the instantaneous brightness, with the variables changed to the ξ, η coordinates and the order of integration with respect to t and η reversed as a matter of convenience, the coefficients become (taking due account of line overlap):

$$A'_{k'l'} = \frac{1}{hw}\int_{-\infty}^{+\infty}\int_{-\frac{h}{2u}-\frac{\eta}{u}}^{\frac{h}{2u}-\frac{\eta}{u}}\int_{-\infty}^{+\infty} S'(\xi,\eta) \exp - 2\pi i\left(\frac{k'\xi}{w} + \frac{l'\eta}{h}\right)$$

$$\exp -2\pi i\left(\frac{k'v}{w} + \frac{l'u}{h}\right)tE(t)d\xi dt d\eta. \quad (6.23)$$

Considering the integral in time, which is, omitting constant terms and substituting for $E(t)$,

$$\int \exp - 2\pi i\left[\frac{(k - k')v}{w} + \frac{(l - l')u}{h}\right]t\, dt,$$

it is evident that the only terms contributing to the coefficients $A'_{k'l'}$ are those for which the relation

$$\frac{kv}{w} + \frac{lu}{h} = \frac{k'v}{w} + \frac{l'u}{h},$$

$$k + \frac{l}{n} = k' + \frac{l'}{n} \quad (6.24)$$

is fulfilled.

One condition satisfying Eq. 6.24 is:

$$k = k',$$
$$l = l'.$$

This gives rise to a direct correspondence between the coefficients in the transmitted picture and received image. The amplitudes of the corresponding coefficients are related as follows:

$$A'_{kl} = A_{kl}Y(k,l)Y'(k,l) \quad (6.25)$$

where, for symmetrical spots:

$$Y'(k,l) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} S'(\xi,\eta) \cos 2\pi\left(\frac{k\xi}{w} + \frac{l\eta}{h}\right) d\xi d\eta. \qquad (6.25a)$$

The factor $Y'(k,l)$ bears the same relation to the receiving spot as $Y(k,l)$ bears to the exploring element. In other words, it is the aperture admittance of the reproducing spot.

No account has been taken in Eq. 6.25 of the transmission characteristic of the communication channel. Since it is a function of frequency, it obviously has an effect on the relative amplitude of corresponding coefficients. The frequency response of the channel may be represented by the function $G(f)$, which, because of the relation between f and k,l, can be expressed as $G(nN[k + l/n])$. Therefore, including this in Eq. 6.25, the coefficients at the transmitted and received pictures are given by:

$$A'_{kl} = A_{kl}G\left(nN\left[k + \frac{l}{n}\right]\right)Y(k,l)\,Y'(k,l).$$

From the standpoint of horizontal resolution the admittance may be treated as a function of the frequency of the signal components. The relation between the coefficients in the transmitted and reproduced picture then is:

$$A'_f = A_f G(f) Y'(f) Y(f).$$

It will be noticed that the spot admittances have the same form as the channel response, and therefore, by making the latter proportional to the reciprocal of the spot admittances, the effect of the decreasing admittance with increasing frequency can be compensated. In practice this compensation is only partial. Compensation of this type can be used for the horizontal frequency components even with an unsymmetrical spot.

The condition $k = k'$, $l = l'$ is not the only one which satisfies Eq. 6.24. Therefore, as well as the desired corresponding Fourier components, some extraneous components will be present in the reproduced picture which do not exist in the transmitted image. These components will be:

$$A'_{k'l'} = A_{kl}G\left(nN\left[k + \frac{l}{n}\right]\right)Y'(k',l')\,Y(k,l),$$

where

$$k + \frac{l}{n} = k' + \frac{l'}{n}, \begin{cases} k \neq k' \\ l \neq l' \end{cases}.$$

These extraneous components give the line structure of the reproduced image. Because of them the vertical resolution is restricted to a step process, and a bright point moving in a vertical direction moves in finite increments and may change in size periodically as it progresses. In addition, these components may cause one or more false patterns to accompany the normal pattern. This is very strikingly demonstrated in the

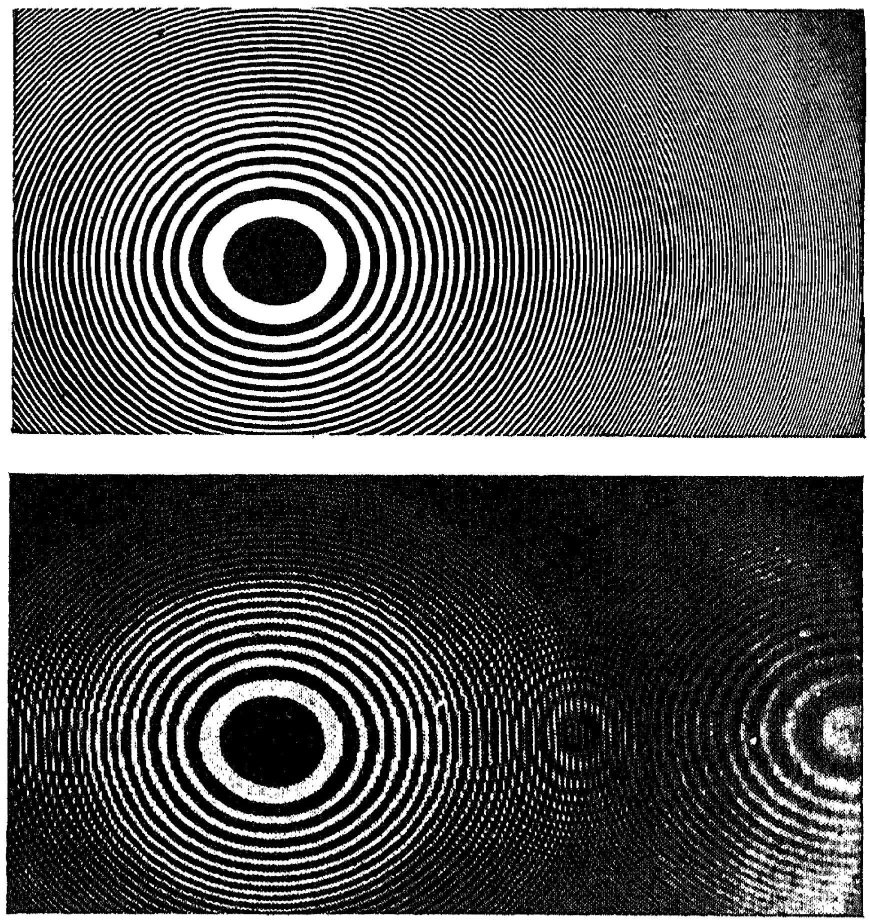


FIG. 6.26.—Spurious Patterns in the Reproduction of a Fresnel Zone Plate. (Courtesy of Bell Telephone Laboratories)

transmitted reproduction of a zone plate shown in Fig. 6.26. The process of generating such false patterns is often graphically, but not quite accurately, referred to as beats between the scanning pattern and the image.

6.8. More about Resolution. One of the most controversial subjects in the field of television relates to the obtaining of optimum resolution. The evidence available at present indicates that the reproduced picture should have approximately equal horizontal and vertical definition. Furthermore, the resolution should be as high as possible, but always subject to the condition that extraneous signals due to the scanning process and to the shape of the channel pass band shall be within the limits of toleration.

As was pointed out in section 6.6, the structure of a 441-line picture is sufficiently fine so as to be unobtrusive when observed from a comfortable viewing distance. Therefore a 441-line scanning pattern has been adopted for use in this country.

The maximum vertical resolution is obtained by means of very small scanning apertures at the transmitter and reproducer. Such a procedure will not give optimum results, however, because of spurious images due to line structure.

The existence of these spurious images was indicated by the extraneous components shown to be present in the received picture. The effect is particularly objectionable if the spot size is smaller than the line separation. For example, if the exploring spot at the transmitter is less than the width of the vertical line separation, the reproduction of a bright line, nearly parallel to the scanning lines, will be a row of separated bright beads, clearly visible even at distances for which the line structure cannot be resolved. This and other spurious effects due to line structure can be minimized by making the distribution of the spot such that the response and brightness at the transmitting and receiving ends, respectively, are uniform over the entire field. A field meeting this requirement will be termed a flat field. Rectangular spots having uniform distribution, whose heights are equal to the line separation, will give a flat field.

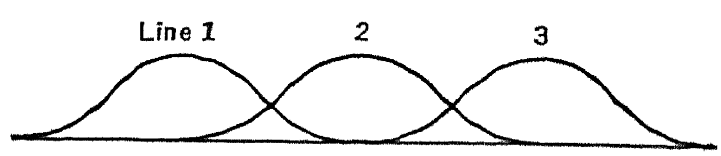


FIG. 6.27.—Overlapping Cosine-Squared Spot Which Yields a Flat Field.

In most cathode-ray systems the spot has radial symmetry, and a distribution such that the brightness or response is maximum in the center. To obtain a flat field in this case it is necessary to overlap the lines. The actual distribution across a scanning line approximates the error function. It also can be represented by a cosine squared distribution with a fair degree of accuracy, this distribution having certain advantages in mathematical manipulation. The minimum size of a spot having this distribution which will give a flat field requires a 50 per cent overlap. The distribution for this type of spot is illustrated in Fig. 6.27.

It is important, therefore, to determine the channel bandwidth which will give a horizontal resolution equal to the vertical resolution obtainable with a 441-line pattern and a cosine squared spot having a 50 per cent overlap.

Before proceeding any further, a satisfactory criterion for resolution in the two directions must be established. This is difficult because of the unusual nature of the picture structure and the little that is known of the visual response to pictures having an anisotropic structure.

The resolution in the vertical direction is necessarily an average because of the line structure. However, if the lines overlap, a condition which favors the reduction of extraneous signal, the vertical resolution becomes more nearly a constant independent of the position of the detail being resolved.

Two of the several methods of evaluating resolution will be mentioned. One of these, due to Wheeler and Loughren,* makes use of the width of confusion, which is defined as the width of a uniform band of light whose brightness is equal to the peak brightness of the reproduction of a very narrow horizontal line, and whose total light flux is equal to that of the reproduction. Fig. 6.28 illustrates the meaning of width of confusion. The narrow rectangular peak (*a*) represents the original image in cross-section, brightness being plotted vertically, while the broad rounded curve (*b*) shows the brightness variation across the reproduced line. On the second curve is drawn a rectangle having equal height and area, whose width is the width of confusion of the reproduction.

When a narrow line used as original is exactly parallel to the scanning lines, the width of the reproduction will depend upon the position of the

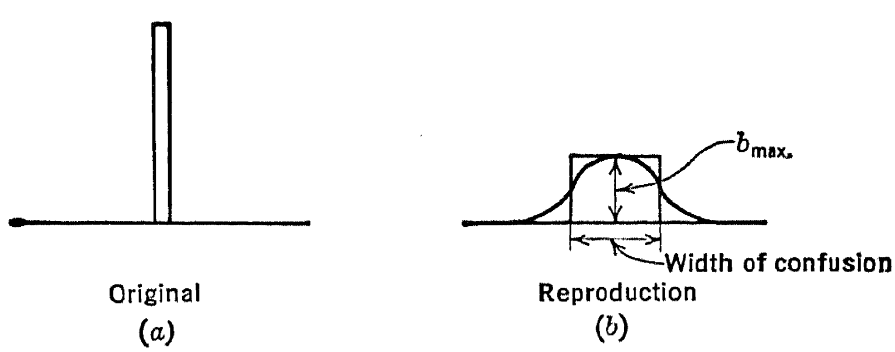


Fig. 6.28.—Width of Confusion.

original. On the other hand, when the line makes a small angle with the scanning lines, the vertical width of confusion will vary over the length of the reproduction. The vertical resolution is the average width of confusion of the reproduction of a narrow line making a small angle to the scanning lines.

If the brightness across the lines making up the scanning pattern at the reproducer varies as the cosine squared, and the sensitivity across the exploring lines at the transmitter varies similarly, assuming a 50 per cent

* See Wheeler and Loughren, reference 6.

overlap of the scanning lines, the average vertical width of confusion can be shown to be 1.414 times the separation between lines.

The definition of width of confusion can also be applied in the horizontal direction. A Fourier transformation can be used to determine the shape of the received image after the transmitted frequency components have been attenuated by the aperture defects of the two scanning spots and the channel limitations. The resolution criterion is then applied to the resulting image. The spectrum of a sharp pulse is essentially constant at all frequencies. The upper frequencies will be attenuated by the admittance of the two spots and by the channel response, so that the reproduced pulse will be the result of a band of frequencies whose amplitudes decrease gradually to zero with increasing frequency, as is shown in Fig. 6.29. It is desired that this band of frequencies be such that the reproduced pulse have a width of confusion equal to $\sqrt{2}$ times the vertical line separation. The theory of Fourier transformation states that the width of confusion is equal to a half period of the nominal cutoff frequency f_c of the pass band, where the nominal cutoff is defined as the upper frequency limit of a frequency band having constant amplitude equal to the low-frequency amplitude of the pass band, and whose area is the same as that of the pass band. The nominal cutoff is indicated for the band-pass characteristics of the entire channel in Fig. 6.29. The nominal cutoff required to give the desired width of confusion must be:

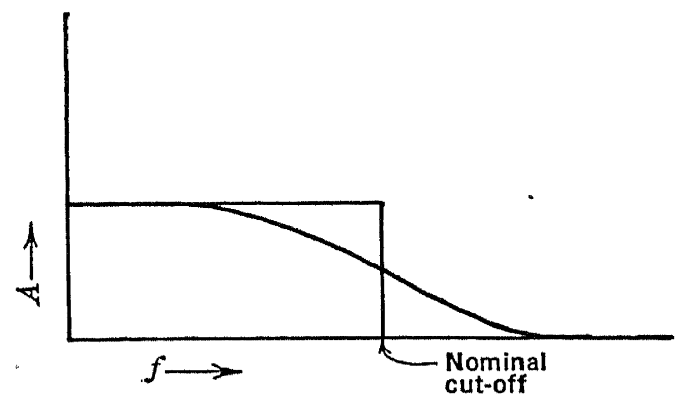


FIG. 6.29.—Attenuation of the High Frequencies by the Communication Channel.

$$f_c = \frac{1}{2\sqrt{2}}\left(\frac{w}{h}\right)n^2N,$$

$$f_c = 2.75 \times 10^6 \text{ cycles},$$

for the number of frames and lines adopted for the standard scanning pattern.

The distribution of the spots having already been specified, their admittance is fixed as shown in Fig. 6.30, together with their nominal cutoff frequency. If the channel has the characteristic shown in the dotted curve (b), having the nominal cutoff of 1.62 f_c and zero response at $2f_c$, the overall pass band has the required characteristic. This condition is

satisfied if the channel is flat to 2.75 megacycles and falls to zero response at 5.5 megacycles. However, reduction of this band to one which is flat to 2.75 megacycles and then gradually decreases to zero at 4.25 megacycles can be effected without too serious loss in horizontal resolution.

A second criterion for vertical resolution is the so-called width of blurring, and is due to Bedford and Fredendall.* Instead of a narrow line, or a point, being used as the original image whose reproduction gives a measure of resolution, an abrupt transition from light to darkness in the vertical direction is employed. The reproduction of such a transition changes from light to dark more gradually than the original, and the extent of this region of change serves as a measure of the resolution. In order to evaluate the width of the region of change, the transition is ap-

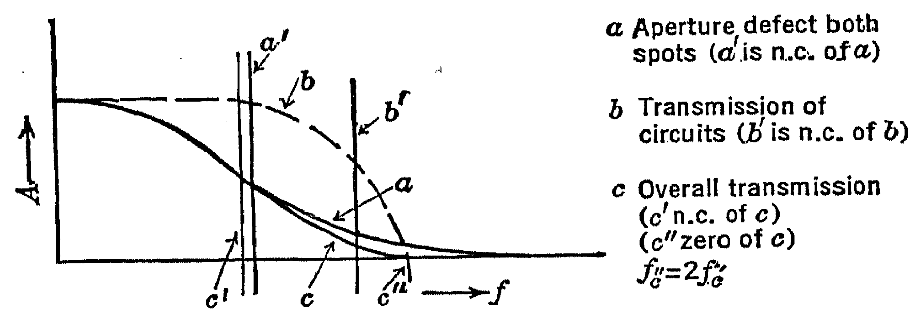


FIG. 6.30—Channel Characteristic as Resultant of Aperture Defect and Circuit Response.

proximated by a linear variation of brightness by a process of averaging. The averaging process fulfills the following two conditions:

1. The integrated light intensity over the transition region is equal to that over the linear variation.
2. The integral of the product of the light intensity and a weighting factor over the transition is equal to the integral of the linear representation of the transition multiplied by the same weighting factor. The weighting factor varies from +1 to −1 over the region covered by the uniform representation of the transition, and outside this region falls linearly to zero on each side over a distance equal to half this interval.

Fig. 6.31 illustrates the actual transition and its representation, together with its weighting curve for a particular reproduction of a vertical edge.

With the same overlap and distribution of scanning pattern described for the previous method of determining resolution, the average region of blurring in the vertical direction is found to be 1.32 times the separation of the scanning lines.

This criterion can also be used to determine the performance in a hori-

* See Bedford and Fredendall, reference 10.

zontal direction. The response of the system as a whole to an abrupt transition from light to dark can be determined by means of Fourier integrals in much the same way as for the pulse cited above. The region of blurring of the reproduction of the transition is then determined by the rules outlined. The proper characteristics for the system, including channel and aperture admittance, are such as result in a region of blurring equal to 1.32 times the line separation. Bedford and Fredendall find that, using a cosine square line distribution, a suitable band shape is one which has constant transmission to 2.77 megacycles and then decreases to zero at about 3.6 megacycles.

It should be pointed out that the pass band of the communication channel cannot be cut off sharply without introducing spurious images

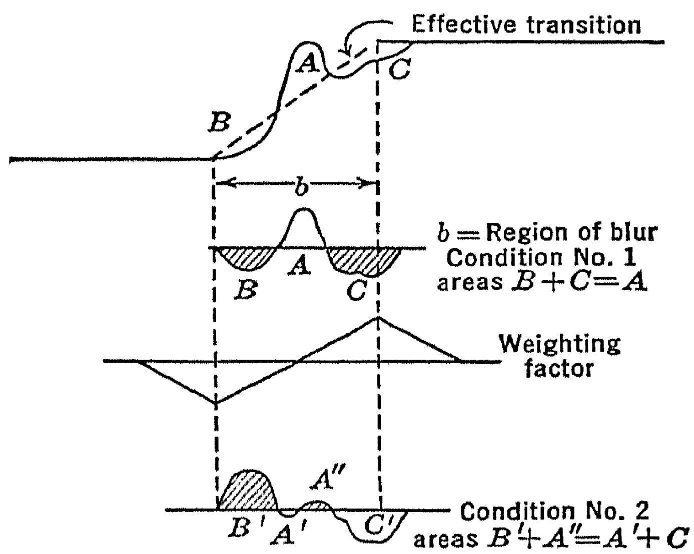


Fig. 6.31.—Determination of Equivalent Blur.

due to transients. The horizontal transients consequent on any abrupt changes from light to dark, or vice versa, in the image being transmitted, if the communication channel has an abrupt cutoff, can easily be shown with the aid of Fourier transformation theory. This effect was observable in the example cited in section 6.6 of a sharp pulse being transmitted through a channel having a rectangular transmission band. The main reproduced pulse in this case was preceded and followed by a series of lesser pulses, so that, if the original image had been a single narrow vertical line, the reproduction would have consisted of a family of lines.

Such transients are also introduced by any marked irregularities in the amplitude-frequency response of the channel. An even more serious source of unwanted transients is phase distortion, which if present even in a small amount may seriously curtail the value of the reproduced picture. Therefore the overall response of the channel, including the aper-

ture defects of both spots, should have an attenuation which increases smoothly with frequency, and should be free from phase distortion (i.e., should have a time delay which is independent of frequency).

The available band is determined by the ultra-high-frequency radio channels which have been allotted to television broadcasting by the Federal Communications Commission, and which are 6 megacycles wide. This 6 megacycles must include the accompanying sound channel, one complete sideband of video signal, and a portion of the second video sideband; therefore the bandwidth which can be used to transmit the picture signal cannot be greater than about 4 to 4.25 megacycles. This, according to the estimates given above, should suffice to give nearly equal horizontal and vertical resolution. The relation between the radio signal and the video signal will be discussed in some detail in Chapter 7 and, except as limiting the channel width, need not have been mentioned at all at this point.

So far it has been tacitly assumed that, except for arbitrary restrictions on the radio channel and certain practical considerations, the bandwidth of the overall channel could be increased indefinitely at no expense of picture quality. This is not entirely true, as it leaves out of consideration statistical electrical fluctuations. In any electrical system whatsoever, the energy of these random fluctuations, which for convenience will be called "noise," is spread through the entire frequency spectrum. As the frequency band transmitted is increased, the noise increases accordingly. If the noise is appreciable compared with the picture signal, it appears in the reproduction as a myriad of constantly changing bright specks. When pronounced, this effect has been likened to a snowstorm in the picture.

Because of this, any increase in bandwidth makes necessary the use of more video signal, or, more exactly, the video signal at every point in the system must be kept above an increased noise level. This is particularly important at the pickup device itself, and also at the radio receiver where the picture is being reproduced. The details of the role played by noise at the pickup device, in the picture amplifier, and at the radio receiver will be taken up in chapters which follow. It suffices to state at this point that the maximum bandwidth which can be used is determined by noise and its relation to the sensitivity of the overall system.

REFERENCES

1. J. W. T. Walsh, "Photometry," Constable, London, 1926.
2. J. C. Wilson, "Television Engineering," Pitman, London, 1937.
3. E. W. Engstrom, "A Study of Television Image Characteristics," Part I, *Proc. I. R. E.*, Vol. 21, pp. 1631–1651, December, 1933.
4. Part II of reference 3, *Proc. I. R. E.*, Vol. 23, pp. 295–310, April, 1935.

5. PIERRE MERTZ and FRANK GRAY, "Theory of Scanning," *Bell Sys. Tech. J.*, Vol. 13, pp. 464-515, July, 1934.
6. H. A. WHEELER and A. V. LOUGHREN, "The Line Structure of Television Images," *Proc. I. R. E.*, Vol. 26, pp. 540-575, May, 1938.
7. A. V. BEDFORD, "A Figure of Merit for Television Performance," *R.C.A. Rev.*, Vol. 3, pp. 36-44, July, 1938.
8. S. W. SEELEY, "Effect of the Receiving Antenna on Television Reception Fidelity," *R.C.A. Review*, Vol. 2, pp. 433-441, April, 1938.
9. I. G. MALOFF, "Gamma and Range in Television, *R.C.A. Rev.*, Vol. 3, pp. 409-417, April, 1939.
10. A. V. BEDFORD and G. L. FREDENDALL, "Relation between Band Width and Number of Lines in Television Picture" (unpublished).
11. R. R. LAW, "Contrast in Kinescopes," *Proc. I. R. E.* Vol. 27, pp. 511-524, Aug. 1939.
12. J. C. WILSON, "Channel Width and Resolving Power in Television Systems," *J. Television Soc.*, Series 2, Vol. 2, pp. 397-419, June, 1938.

CHAPTER 7

THE TRANSMISSION AND REPRODUCTION OF HIGH-DEFINITION PICTURES

The preceding chapter put forward certain requirements that must be met by an idealized skeleton television system in order to produce a high-definition picture. These requirements are not out of keeping with what can be physically realized with the resources available. The purpose here is to outline the actual embodiment of a high-definition system, or more exactly, high-definition systems in general, for the discussion will center chiefly around those elements which are common to most systems.

Briefly, the system may be subdivided into the following elements:

1. Pickup device.
2. Video amplifier.
3. Radio transmitter.
4. Radio receiver.
5. Receiver amplifier.
6. Viewing device.
7. Synchronizing equipment.

Schematically they can be represented as shown in Fig. 7.1. Because the system functions as an entity, this subdivision is not without its disadvantages. However, for the sake of simplicity and clarity it has been adopted.

7.1. The Pickup Device. The pickup device is the element in the system which converts the light image into the train of electrical impulses making up the video signal. As has already been explained, this involves scanning the picture area with a suitable exploring element. There are a number of ways in which the conversion of a light image into the video signal may be accomplished. The means for this conversion fall into one of two categories, namely, electronic and mechanical systems.

The two principal electronic pickup devices are the Iconoscope and the Dissector Tube. The former functions by storing on a light-sensitive mosaic a charge image which is a replica of the optical picture of the scene being televised. The charge is removed point by point from the mosaic by a fine pencil of electrons which scans the stored image. The released charge forms the picture signal. The Dissector Tube makes use

of a quite different principle. An electron image is formed by focusing the electrons from a photocathode on which the picture being televised is projected. This electron image is moved across a small opening so that it is scanned by this aperture. The current entering the aperture is amplified by a secondary-emission multiplier and constitutes the video signal. Both these tubes depend for their scanning upon the deflection of an electron stream. The deflection is accomplished by means of suitably directed and varied electrostatic or magnetic fields. The voltage or current producing the fields is generated by a deflection generator, which in turn is controlled by synchronizing equipment.

In general, mechanical pickup devices do not lend themselves to high-definition television work. However, they have been used successfully in connection with the transmission of moving pictures. For this a Nipkow

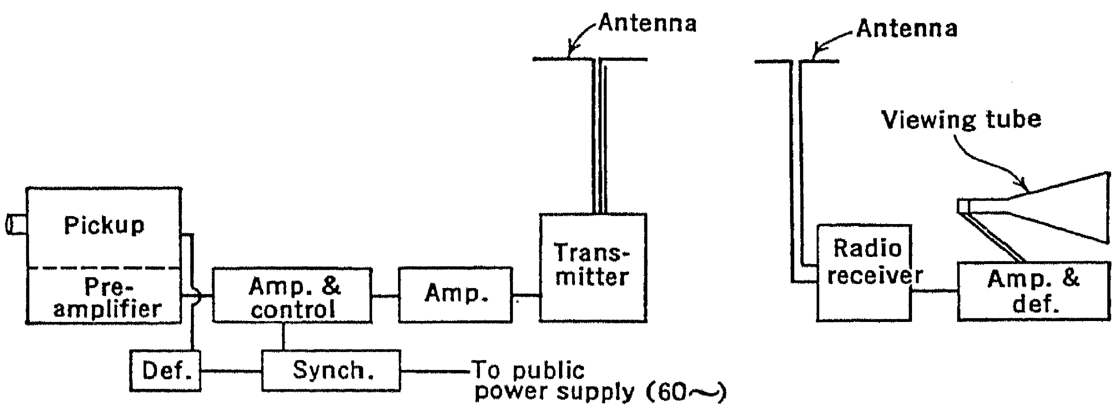


FIG. 7.1.—Elements of a Television System.

disk, often running in vacuum, rotates in the plane of the image, apertures in the disk moving across the image to produce the scanning pattern. Behind the disk is the light-sensitive element, usually in the form of a phototube-multiplier, which converts the light passing through the apertures into the video signal. The rotation of the disk is controlled by a signal from the synchronizing equipment.

These pickup devices will be described in greater detail in the next chapter; the above brief account merely indicates the role this equipment plays in the system as a whole.

7.2. The Amplifiers. The signal output from any one of the above pickup devices usually represents only a minute amount of power—not more than a few thousandths of a microwatt—and a voltage level of the order of a few millivolts. Before this signal can be used to modulate a broadcast transmitter of even moderate power, it must be amplified to a level of several hundred watts. Thus the power gain of the chain of amplifiers raising the signal level is of the order of 10^{12}. Allowing a power gain of four per amplifier stage, a minimum of twenty stages is required.

Actually two or three times this number are used, because at many points in the system it is necessary to make up for attenuation losses in coaxial cables, losses incident in controlling and shaping the signal, etc.

This amplifier, like those employed for audio amplification purposes, makes use of thermionic vacuum tubes. However, because of the much more exacting requirements imposed by the amplification of the video signal, the means used to cascade the tubes are somewhat different. A discussion of the design of this type of amplifier will be deferred until Chapter 14 and only the general purpose and nature of the video amplifier will be considered in this section.

The frequency band required to give the desired resolution was shown to be between 4 and 4½ megacycles wide, and the amplifiers must have a uniform gain and constant time delay for this entire range of frequencies. The relatively small gain that can be obtained in each stage is due to the very wide band of frequencies which must be amplified without distortion. An exception to the requirement of uniform gain occurs when it is necessary to correct for the aperture defect of the scanning spots, in which case the gain at high frequencies is made greater than that at low frequencies in order to compensate for the reduction in amplitude of the former due to the finite spot size.

The signal which is generated by the pickup device does not contain the complete information necessary for the reconstruction of the picture. In addition to voltage variations representing the light and shade in the picture, it is necessary to furnish the information required to insure synchronism between the scanning pattern at the transmitter and the receiver. This is supplied in the form of properly shaped pulses, timed so that they occur after the end of each scanning line and before the beginning of the next. Similarly a pulse or series of pulses is added between frames or field repetitions to insure vertical synchronization.

The synchronizing impulses are added to the picture signal in the amplifier chain. This operation is performed by rendering the amplifier inactive, ahead of the point of injection, for a period equal to or slightly in excess of that occupied by the actual synchronizing signal, and injecting the impulse during this interval. The purpose of this "blanking" process is to prevent any possible signal from the pickup device from interfering with the synchronizing signal. During the blanking period the voltage level in the amplifier is brought to a value which corresponds to maximum white or, what is more usual, maximum black of the picture, thus providing a reference point from which the average illumination in the picture is indicated. If the pickup device is essentially an a-c device indicating only the variation in light and dark over the picture, but not giving information as to the average illumination, as, e.g., an Iconoscope,

a special phototube and d-c amplifier may be necessary to establish this value. The various steps in the formation of the complete signal are indicated in Fig. 7.2.

For convenience, the amplifier at the transmitter may be considered as consisting of three parts. At the pickup device there is a preamplifier which receives the very weak signal, usually of the order of a millivolt or at most a few millivolts, and raises it to a level of perhaps a tenth of a volt. This amplifier is made to have as low a noise level as possible, because the input at this point is not greatly in excess of the statistical fluctuations in the necessary coupling resistors and vacuum tubes. As will be shown later, in order to obtain a maximum signal-to-noise ratio, it is necessary to make the gain of the first few stages as high as possible. This

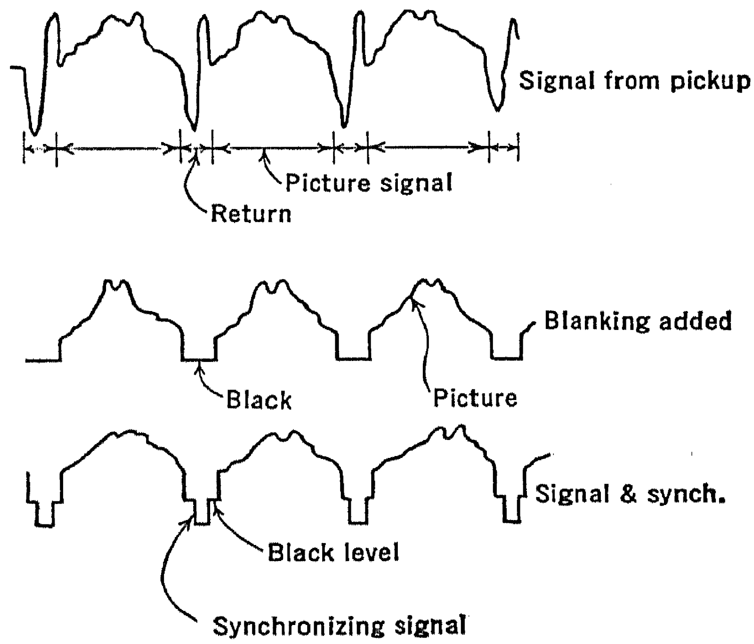


FIG. 7.2—Formation of the Complete Video Signal.

brings in the need of rather complicated coupling arrangements to maintain constancy of frequency response.

The output of the preamplifier is sufficient to drive a cable (usually a low-loss coaxial) which connects it with a second amplifier chain, variously called the monitoring amplifier or the control amplifier. This unit raises the signal to a level of 1 to 10 volts, and it is in this amplifier that the synchronizing signal is added. Also the gain, the d-c level, etc., are controlled in this amplifier, and any special signals which may be necessary are added.

The output of this amplifier is fed to the modulation amplifier which raises the signal level to that required by the modulator of the radio transmitter.

This subdivision of the amplifier chain into three, and only three,

physical units is, of course, a simplification which does not exist in most practical transmitters, because the physical layout makes more than three units advisable. Functionally these three units can usually be identified.

7.3. The Video Transmitter. The design problem presented by the radio transmitter to be used for television broadcasting is one of the most formidable encountered in the whole television system. This is because of the tremendous bandwidth of frequencies which must be radiated if high definition is to be attained.

Like a sound transmitter, the transmitter in question is made up of a carrier generator which produces a high-frequency oscillation of constant amplitude, and a modulator which varies the amplitude of this radio-

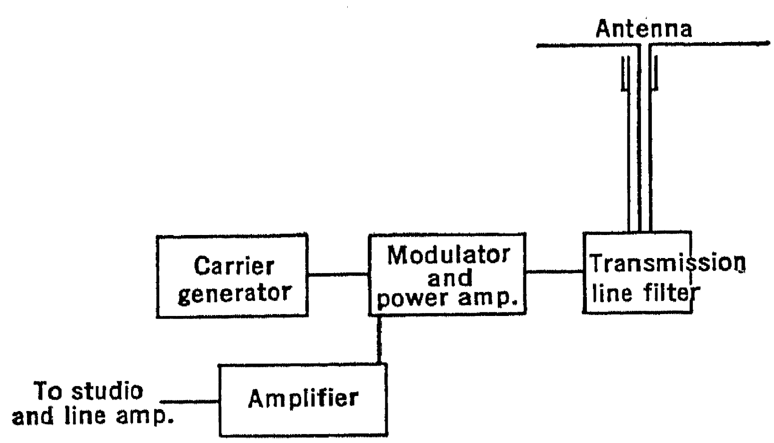


Fig. 7.3.—Block Diagram of a Television Transmitter.

frequency carrier in such a way that it is at every instant proportional to the signal applied to the modulator, in this case the video signal. The modulated radio frequency is then fed to the transmitting antenna, from which it is radiated. This is shown schematically in Fig. 7.3.

Obviously, since the amplitude of the radio-frequency output of the modulator is not constant, the signal no longer consists of a single frequency as does the unmodulated carrier. The existence of these several frequencies can be demonstrated by a brief consideration of the process of modulation. For this purpose let the carrier frequency be f and the modulating frequency p. A factor must also be included which represents the amount by which the amplitude of the carrier is varied, and which is proportional to the amplitude of the modulating frequency. This factor k, when expressed as a percentage, is known as the per cent modulation. It varies from zero, when the amplitude of the modulating frequency is zero, to unity (or 100 per cent) when the amplitude is equal to that of the carrier. In terms of these symbols the modulated signal is:

$$A(1 + k \cos 2\pi pt) \cos 2\pi ft,$$

which by trigonometry can be transformed to:

$$A\left\{\cos 2\pi ft + \frac{k}{2}\cos 2\pi(f+p)t + \frac{k}{2}\cos 2\pi(f-p)t\right\}. \tag{7.1}$$

As was shown in the preceding chapter, the video frequency p may have any value from zero (approximately) to 4 or 4½ megacycles. Therefore the band of frequencies which must be transmitted is $2\,p_{\text{max.}}$ cycles wide —i.e., 8 or 9 megacycles wide—with its center at the carrier frequency. These additional frequencies above and below the carrier are known as the upper and lower sidebands, respectively.

In order to accommodate this wide band of frequencies the carrier frequency itself must be very high—i.e., in the ultra-short-wave region—if more than one broadcasting station is to function at any one time. Actually frequencies in the neighborhood of 50 megacycles are used for television broadcasting, while frequencies between 150 and 300 megacycles, or higher, are used for television relay purposes. Recently the

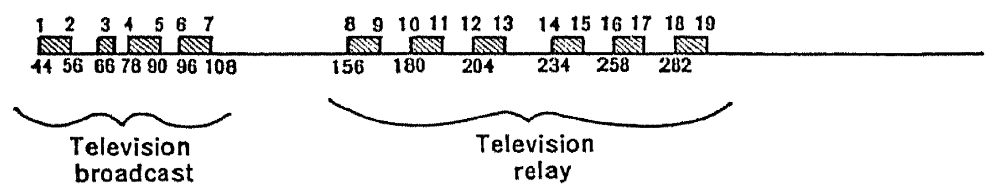


Fig. 7.4.—Distribution of Television Channels in the Ultra-High-Frequency Region.

Federal Communications Commission made a tentative frequency allotment of the ultra-short-wave region, assigning certain frequencies for television use. A chart showing the subdivision is given in Fig. 7.4. The television broadcast frequencies will probably be those lying between 40 and 85 megacycles. The reason for confining television broadcasting to the region of lower frequencies of the ultra-short-wave band is the technical difficulty of generating wide-band radio signals of high power at extremely short wavelengths due to the limitation of present-day power amplifiers. These difficulties will be reviewed in some detail in Chapter 16.

It will be noticed in Fig. 7.4 that the channel width assigned for each television transmitter is 6 megacycles. In view of the 8 or 9 megacycles indicated by Eq. 7.1 it would seem that this would seriously limit the definition of the pictures which could be broadcast. Actually this is not so, because it can be shown that all the information in the video signal is present if the frequencies transmitted include the carrier and all frequencies above the carrier, that is, a band from f to $f + p_{\text{max.}}$ Similarly, if the carrier and all frequencies below it are transmitted, the signal contains the full information. Therefore the bandwidth necessary needs to be only slightly greater than the 4 to 4½ megacycles. For practical rea-

sons connected with the means adopted to eliminate the unwanted sideband, the frequencies of this sideband which are close to the carrier are included in the signal. Furthermore, the sound which supplements the picture is also transmitted in this channel, albeit with a separate carrier. These signals fill the 6-megacycle band, as shown in Fig. 7.5.

The propagation characteristics of the ultra-short-wave signal are somewhat different from those of broadcast or ordinary short-wave signals. This difference arises from the fact that this radiation has much more in common with light than the lower frequencies have. The useful range of ultra-high-frequency transmission is essentially determined by the horizon; in other words, the radiation does not follow the curvature of the earth's surface. Because of this the service area of any one transmitter is limited to a radius of 25 to 50 miles for practical transmitting antenna heights. This is not entirely without its advantages, inasmuch

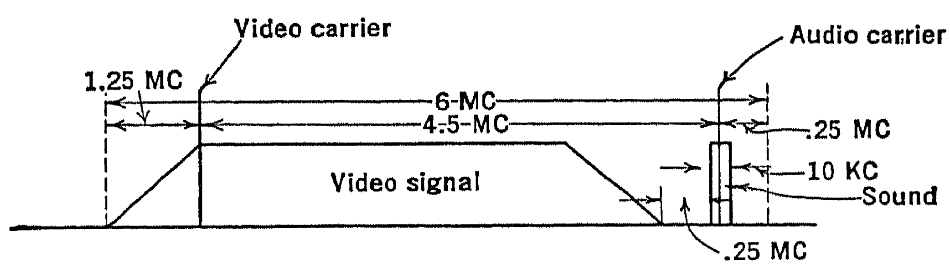


FIG. 7.5.—Utilization of the 6-Megacycle Channel.

as it avoids all difficulties from reflections at the various ion layers in the upper atmosphere which are responsible for the fading encountered at the longer wavelengths. If the properties of ultra-high-frequency radiation resembled those of the lower frequencies, fading would be so severe because of the great bandwidth that the quality of pictures coming from distant points would be seriously impaired. Furthermore, these same unusable distant signals would interfere with local reception.

Owing to the quasi-optical properties of this radiation, it can be very easily directed by the antenna arrangement. The signal can therefore be placed where it is of most value; for example, it can, by means of a suitable transmitting antenna, be confined to a horizontal plane, so that the energy is not wasted in the vertical direction where it cannot be used. Similarly the receiving antenna can be made directive and consequently insensitive to radiation coming from any direction other than that of the transmitter.

7.4. Single-Sideband Transmission and Reception. It was stated, without proof, that a carrier and all the sideband frequencies either above or below the carrier carry the full information contained in the video

signal. Since there is a one-to-one correspondence between the video frequencies and the different frequencies of the sideband and carrier, and also a proportionality between the sideband amplitudes and video-frequency amplitudes, this statement does not seem unreasonable.

It is not immediately obvious that the video signal can be obtained from a carrier and single sideband by the ordinary processes of detection. Therefore the next consideration will be that of the action of a detector on the single-sideband signal. As has already been shown, a carrier f_0 modulated with a single frequency p leads to a signal having the following form:

$$e = A\left(\cos 2\pi f_0 t + \frac{k}{2}\cos 2\pi(f_0 + p)t + \frac{k}{2}\cos 2\pi(f_0 - p)t\right). \quad (7.2)$$

At the receiver, since the entering signal passes through various band-pass filters, there is an attenuation which is not independent of frequency. The attenuation is a complex quantity, and affects both amplitude and phase. If the amplitude factor is $B(f)$ and the phase shift $\phi(f)$, the signal reaching the detector is:

$$e = A\left[B(f_0)\cos(2\pi f_0 t + \phi(f_0)) + \frac{kB(f_0 + p)}{2}\cos(2\pi(f_0 + p)t + \phi(f_0 + p)) + \frac{kB(f_0 - p)}{2}\cos(2\pi(f_0 - p)t + \phi(f_0 - p))\right]. \quad (7.3)$$

When the signal is rectified by a linear detector and the output put through a low-pass filter to eliminate frequencies much above the maximum video-frequency $p_{\max.}$, the resulting signal is the envelope of this expression. If Eq. 7.3. is written in the form:

$$e = V_s \cos(2\pi f_0 t + \phi(f_0)),$$

the factor V_s is for all practical purposes the envelope.*

Therefore the detector output has the form:

$$V_s = A\Big[B^2(f_0) + \frac{k^2}{4}\{B^2(f_0 + p) + B^2(f_0 - p) + 2B(f_0 + p)B(f_0 - p)\cos(4\pi pt + \phi(f_0 + p) - \phi(f_0 - p))\} + kB(f_0)\{B(f_0 + p)\cos(2\pi pt + \phi(f_0 + p) - \phi(f_0)) + B(f_0 - p)\cos(2\pi pt + \phi(f_0) - \phi(f_0 - p))\}\Big]^{\frac{1}{2}}. \quad (7.4)$$

* Actually the envelope should be determined by differentiating Eq. 7.3 with respect to t and setting the result equal to zero to determine the extremes. Except for changes in radio-frequency amplitudes, much more rapid than can actually occur because of video bandwidth limitation, the results are the same.

If, as they should be in a well-designed filter system for double-sideband operation, the following conditions are fulfilled:

$$B(f_0 + p) = B(f_0 - p) = B(p),$$
$$\phi(f_0 + p) - \phi(f_0) = \phi(f_0) - \phi(f_0 - p) = \theta(p),$$

Eq. 7.4 reduces to:

$$V_s = A\left[B^2(f_0) + \frac{k^2B(p)^2}{2}\{1 + \cos(4\pi pt + 2\theta)\} + 2kB(p)B(f_0)\cos(2\pi pt + \theta)\right]^{\frac{1}{2}} \quad (7.5)$$

or

$$V_s = A[B(f_0) + kB(p)\cos(2\pi pt + \theta)]. \quad (7.5a)$$

This equation shows that no distortion is introduced by the detector.

Single-sideband operation can be represented with Eq. 7.4 by placing either one of the sideband attenuation factors $B(f_0 + p)$ or $B(f_0 - p)$ equal to zero.* As far as the calculation is concerned no distinction is necessary between the suppression of the sideband at the transmitter or at the receiver.

Eq. 7.4 being written with $B(f_0 - p) = 0$, it becomes:

$$V_s = A\left[B^2(f_0) + \frac{k^2B^2(f_0 + p)}{4} + kB(f_0)B(f_0 + p)\cos(2\pi pt + \phi(f_0 + p) - \phi(f_0))\right]^{\frac{1}{2}}. \quad (7.6)$$

It is immediately evident that if the modulation factor k is large this equation represents a distorted output. If k is small, the equation can be expanded by the binomial theorem, giving:

$$\begin{aligned} V_s &= A\left[B(f_0) + \frac{kB(f_0 + p)}{2}\cos(2\pi pt + \phi(f_0 + p) - \phi(f_0))\right] \\ &= A\left[B(f_0) + \frac{kB(p)}{2}\cos(2\pi pt + \theta)\right], \end{aligned} \quad (7.7)$$

using the notation of Eq. 7.5*a*. This equation has the same form as Eq. 7.5*a* and indicates that the detector introduces no distortion.

The fidelity of the response can be seen to depend upon the attenuation characteristic of the filter in both Eq. 7.5*a* and 7.7. In order to have a response independent of the video frequency p the attenuation factor $B(p)$ must be constant. Furthermore, unless the phase factor θ is a linear

* See Poch and Epstein, reference 6.

function of p, the time delay will depend upon frequency with consequent distortion of the reproduced picture.

If one sideband, for example, the lower, is only partially suppressed, that is, if frequencies from $f_0 + p_{\max.}$ to $f_0 - p_1$ are received at the detector and those from $f_0 - p_1$ to $f_0 - p_{\max.}$ are rejected, the resulting video signal will have twice the amplitude over the range 0 to p_1 as from p_1 to $p_{\max.}$ This is illustrated in Fig. 7.6*b*.

On the other hand, if the pass band includes the upper sideband from $f_0 + p_{\max.}$ to $f_0 + p_1$, uniformly, then tapers linearly to zero between $f_0 + p_1$ to $f_0 - p_1$, as shown in Fig. 7.6*c*, the output will be undistorted, at least for a low percentage modulation.

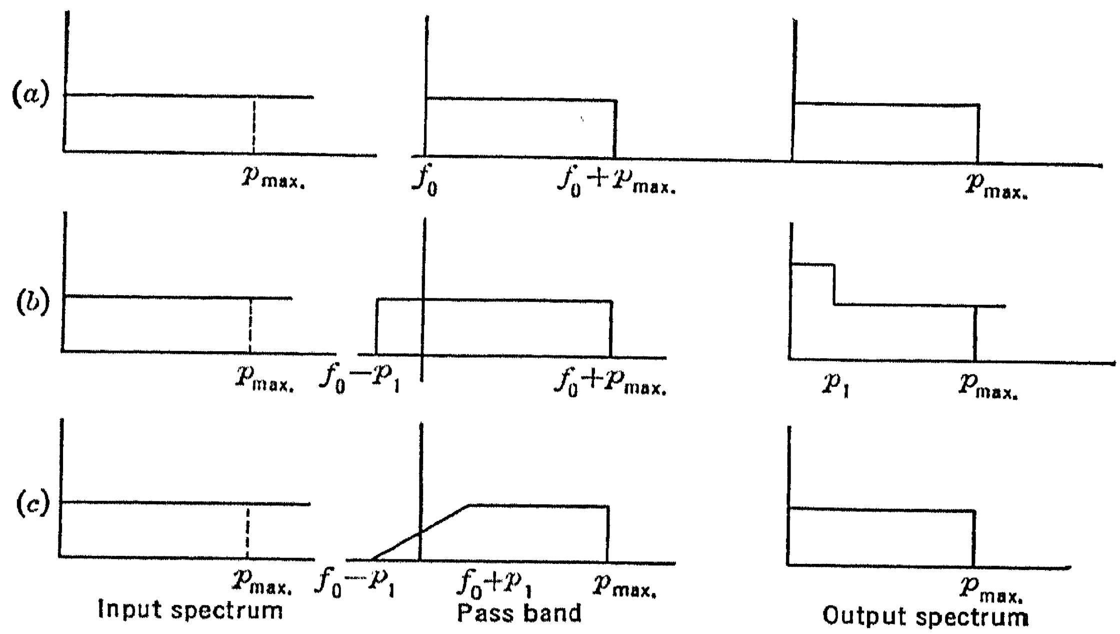


FIG. 7.6.—Effect of Attenuating a Sideband on the Character of the Transmitted Frequency Spectrum.

As is apparent from Eq. 7.6, truly undistorted linear detection is possible only when the percentage modulation is small. For a modulation fraction which is not negligible compared with unity, higher-order terms in the expansion of Eq. 7.6 cannot be dropped. Under these conditions the detector output consists not only of the original video frequency but also of harmonics of it. At 100 per cent modulation the second harmonic is around 15 per cent and the third harmonic from 3 to 5 per cent. In audio work this amount of distortion would be very serious. However, the reconstructed picture is very insensitive to distortion of this nature, so that the presence of this amount does no visible harm.

So far the discussion has been limited to the behavior of the system towards a single frequency. Of greater interest is the response of the two types of systems to a square pulse. This requires the use of Fourier integrals.

The signal whose response is to be investigated, shown in Fig. 7.7a, consists of a square impulse whose duration is 2τ. Its spectrum can be expressed as:

$$G(\omega) = k\frac{\sin \omega\tau}{\omega}, \tag{7.8}$$

where ω is $2\pi p$, p being the video frequency. Hence the Fourier integral describing the pulse is:

$$\frac{2k}{\pi}\int_0^\infty \frac{\sin \omega\tau}{\omega} \cos \omega t d\omega. \tag{7.8a}$$

This pulse modulates a carrier whose frequency is $f_0 = \nu/2\pi$, giving rise to a signal having the following form:

$$\cos (\nu t + \phi) + \frac{k}{\pi}\int_0^\infty \frac{\sin \omega\tau}{\omega} \{\cos [(\nu + \omega)t + \phi] + \cos [(\nu - \omega)t + \phi]\} d\omega. \tag{7.9}$$

This represents a carrier and both upper and lower sidebands extending over an infinite range on either side of this carrier. In the course of transmission and reception, the radio-frequency signal passes through circuit elements which act as filters, limiting the extent of the sidebands before it reaches the detector. Before it is possible to determine the shape of the video pulse leaving the detector, it is necessary to make certain assumptions about the characteristics of the pass band. Three conditions will be considered; they are illustrated in Fig. 7.7. In the last two the pass bands have equal widths, one passing frequencies from $f_0 - p$ to $f_0 + p$, the other from f_0 to $f_0 + 2p$. Over these bands the attenuation factor $B(f)$ is constant, being zero for all other frequencies. The phase factor will be assumed to be a linear function of frequency $\phi(f) = \alpha p + \beta$. The first filter admits both sidebands, having twice the width of the others, that is, it passes the range of frequencies from $f_0 - 2p$ to $f_0 + 2p$.

The detector output after the signal has been attenuated by the band-pass filter is determined, as before, from the envelope of the modulated signal. The envelope is obtained by factoring out the carrier frequency term by the aid of the trigonometric transformation,

$$\cos (\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta.$$

For the two filters passing double-sideband signals, the detector output is

$$V_s = 1 + \frac{2k}{\pi}\int_0^{\Omega} \frac{\sin \omega\tau}{\omega} \cos \omega t' d\omega$$
$$= 1 + \frac{k}{\pi}(Si[\Omega(t' + \tau)] - Si[\Omega(t' - \tau)]),^* \tag{7.10}$$

where $2\Omega = 4\pi p_{\max}$. is the total width of the pass band.

With single-sideband reception the results are not quite so simple. The signal reaching the detector is:

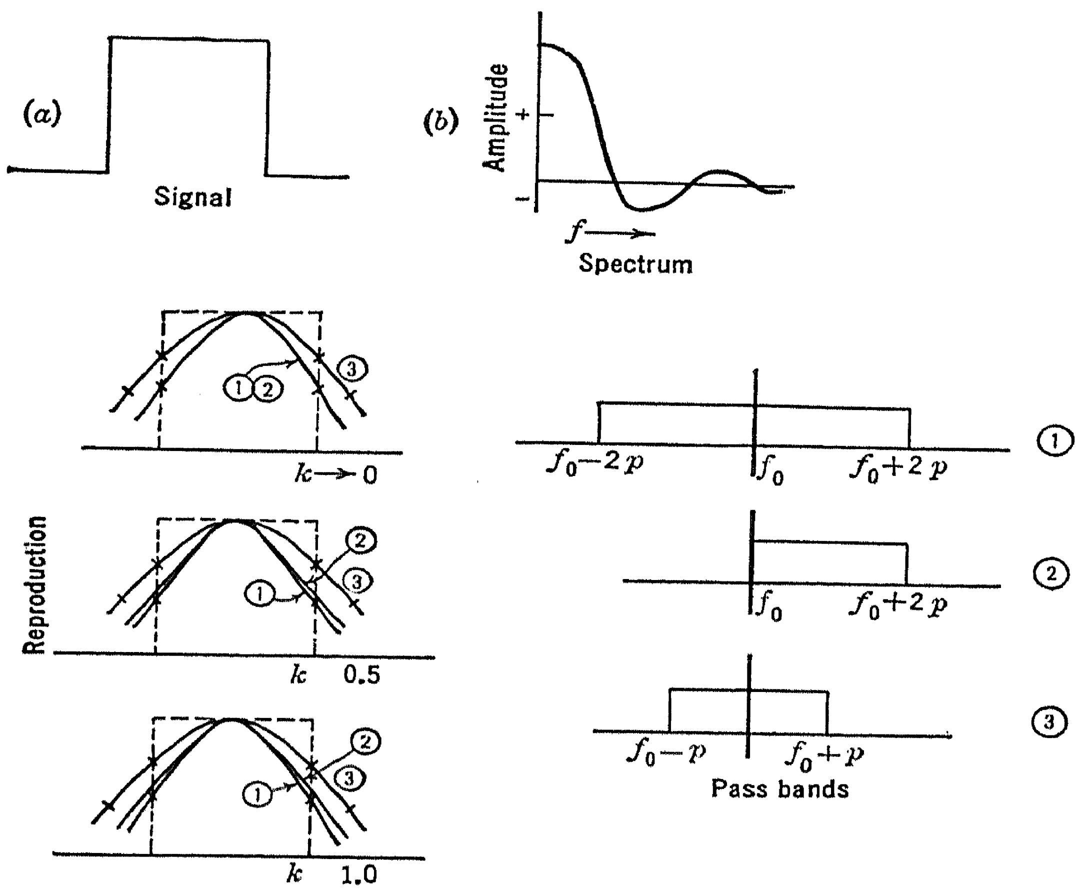


FIG. 7.7.—Comparison of Single- and Double-Sideband Transmission of a Square Pulse.

$$\cos(\nu t + \phi_0) + \frac{k}{\pi}\int_0^{2\Omega} \frac{\sin \omega\tau}{\omega} (\cos(\nu + \omega)t + \phi) d\omega \tag{7.11}$$

or

$$\cos(\nu t + \phi_0) + \frac{k}{\pi}\left(\int_0^{2\Omega} \frac{\sin \omega\tau}{\omega} \cos \omega t' d\omega\right) \cos(\nu t + \phi_0)$$

* $Si(x)$ is the "sine integral" of x

$$Si(x) = \int_0^x \frac{\sin y}{y} dy$$

and $Ci(x)$ is the "cosine integral" of x

$$Ci(x) = \ln x - \int_0^x \frac{1 - \cos y}{y} dy + c$$

$$-\frac{k}{\pi}\left(\int_0^{2\Omega}\frac{\sin\omega\tau}{\omega}\sin\omega t' d\omega\right)\sin(\nu t+\phi_0)$$

$$=\left[1+\frac{k}{2\pi}Si2\Omega(t'+\tau)-\frac{k}{2\pi}Si2\Omega(t'-\tau)\right]\cos(\nu t+\phi_0)$$
$$+\left[\frac{k}{2\pi}Ci2\Omega(t'+\tau)-\frac{k}{2\pi}Ci2\Omega(t'-\tau)-\frac{k}{2\pi}\ln\left(\frac{t'+\tau}{t'-\tau}\right)\right]$$
$$\sin(\nu t+\phi_0), \qquad (7.12)$$

where $t' = t + \alpha/2\pi$ and $\phi_0 = \beta$. This signal has an envelope which is given by:

$$V_s=\sqrt{\left[1+\frac{k}{2\pi}Si2\Omega(t'+\tau)-\frac{k}{2\pi}Si2\Omega(t'-\tau)\right]^2 +\left[\frac{k}{2\pi}\left\{Ci2\Omega(t'+\tau)-Ci2\Omega(t'-\tau)-\ln\left(\frac{t'+\tau}{t'-\tau}\right)\right\}\right]^2}, \qquad (7.13)$$

and the phase shift:

$$\theta=\pm\tan^{-1}\frac{Ci[2\Omega(t'+\tau)]-Ci[2\Omega(t'-\tau)]-\ln\left(\dfrac{t'+\tau}{t'-\tau}\right)}{\dfrac{2\pi}{k}+Si[2\Omega(t'+\tau)]-Si[2\Omega(t'-\tau)]}. \qquad (7.13a)$$

The responses corresponding to the pulses passing through the three pass bands:

(1) $2\Omega = 8\pi p$ — double sideband,
(2) $2\Omega = 4\pi p$ — single sideband,
(3) $2\Omega = 4\pi p$ — double sideband,

are shown for k small, $k = 0.5$, and $k = 1.0$ in Fig. 7.7. It will be seen that even for the worst case, that of $k = 1$, the reproduction for single-side-

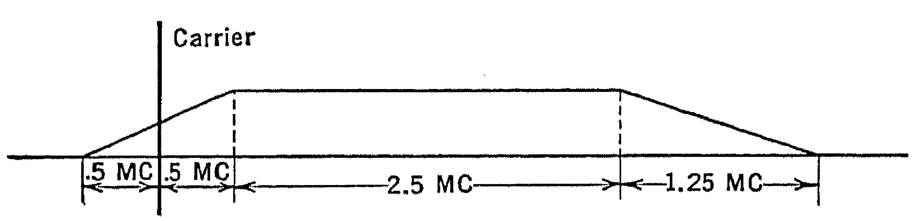


FIG. 7.8.—Pass Band Characteristic of a Receiver Meeting High Definition Requirements.

band operation does not differ greatly from that obtained with double-sideband operation using twice the overall bandwidth. For square law

detection these two coincide even more closely, for all modulations. It is interesting to note that for a step function with single-sideband transmission, where the lower sideband is cut off sharply at the carrier frequency, the envelope does not converge. If, however, the pass band tapers off gradually through the region of the carrier, as shown in Fig. 7.8, this is no longer true.

The conclusions which can be reached from this brief introduction to single-sideband operation are that if the cutoff of the suppressed sideband is not too sharp, and is symmetrical about the carrier, a faithful reproduction of the original video signal can be obtained with linear detection, and that the effective definition will be nearly that of double-sideband operation using twice the overall radio-frequency bandwidth.

A diagram illustrating the ultra-high-frequency spectrum of a single-sideband transmitter was given in Fig. 7.5. The lower sideband only is suppressed, the gradual cutoff beginning at carrier frequency. If this form of signal reached the final detector, there would be a distortion of the picture in the form of accentuated low frequency. Therefore the band-pass characteristics of the radio- and intermediate-frequency stages of the receiver are so adjusted as to give a cutoff which tapers off gradually through the region of the carrier, as shown in Fig. 7.8. With this characteristic the requirements necessary for faithful reproduction with one sideband are met. The reasons for not shaping the transmitter characteristics to give the spectrum needed at the detector are purely practical and connected with the difficulty of designing and constructing a filter capable of handling the large amounts of power, yet having the required attenuation and phase properties.

7.5. Radio Receiver. The signal received by the antenna is conveyed to the radio receiver by some form of balanced line. This line may merely be a twisted pair, if the location of the antenna is such that a relatively short line can be used, and if the signal strength is high. On the other hand, if the feed line is long, and particularly if the field strength at the antenna is small, it may be necessary to use a two-wire, air-insulated transmission line or even a pair of coaxial cables.

The receiving equipment itself provides for converting the radio-frequency signal into a video signal and amplifying it sufficiently so that it can be used to operate the viewing device. It also separates the synchronizing impulses from the picture signal, employing the former to control the scanning pattern of the viewing device.

As with the conventional radio receiver, several alternative types of television receivers are possible. The two most frequently used employ superheterodyne and tuned radio-frequency circuits.

A tuned radio-frequency receiver has five or six tuned stages of ultra-

high-frequency amplification, a detector, and several stages of video amplification. Because of the difficulty of accurately tuning the many radio-frequency stages, it is rarely used where reception from more than one transmitter is desired. Fig. 7.9 shows in block diagram this type of receiver.

The superheterodyne receiver is made up of a radio-frequency element, a first detector and oscillator, an intermediate amplifier, and a second detector followed by a video amplifier. The radio-frequency element may merely be a broadly tuned circuit whose chief function is to match the antenna feeder to the first detector circuit, or it may be more complicated, containing, for example, a stage of radio-frequency amplification. The radio frequency leaving this first stage is mixed with the output from the local oscillator and rectified by the first detector, giving rise to the intermediate frequency. The mixing is merely an addition of the incoming

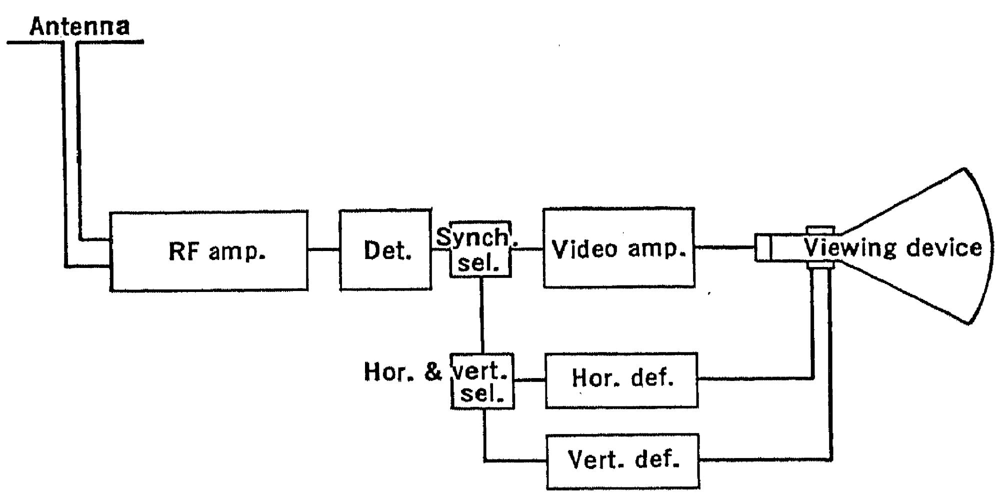


Fig. 7.9.—Television Receiver Employing Radio-Frequency Amplification.

signal and the locally generated frequency, or a modulation of one by the other, producing a beat effect between these two waves. For example, if the incoming signal consists of the carrier alone, and so has the form $A_0 \cos 2\pi f_0 t$, and the local wave is $A_1 \cos 2\pi f_1 t$, the envelope of the resultant wave will be:

$$\sqrt{A_0^2 + A_1^2 + 2A_0A_1 \cos 2\pi(f_0 - f_1)t}. \tag{7.14}$$

The sideband frequencies which are present beat with the locally generated oscillation in a similar way. When the mixed signal is then rectified by an ideal detector, the output contains an intermediate carrier whose frequency is $(f_0 - f_1)$ and sidebands associated with it which correspond exactly to the sidebands present in the original signal. The intermediate carrier and its sideband are amplified by the intermediate amplifier, the other frequencies present in the output of the detector being removed by

the band-pass characteristics of the interstage coupling of this amplifier. Rectification of the mixed signal by a linear detector also leads to the intermediate carrier and sidebands, but there are present, in addition, terms resulting from the addition of various sideband frequencies. However, if the amplitude of the locally generated oscillation is large compared with the incoming signal, the amplitude of the unwanted combinations of sideband frequencies will be small compared with those of the fundamentals. For practical reasons it is common to use a linear detector and a powerful local oscillator rather than a square law detector.

After amplification, the intermediate carrier and sideband are again rectified, this time by the second detector which is also usually a linear detector. The output of this element is a signal which corresponds exactly to the video signal generated by the pickup device together with synchronizing impulses and any correcting signals inserted at the moni-

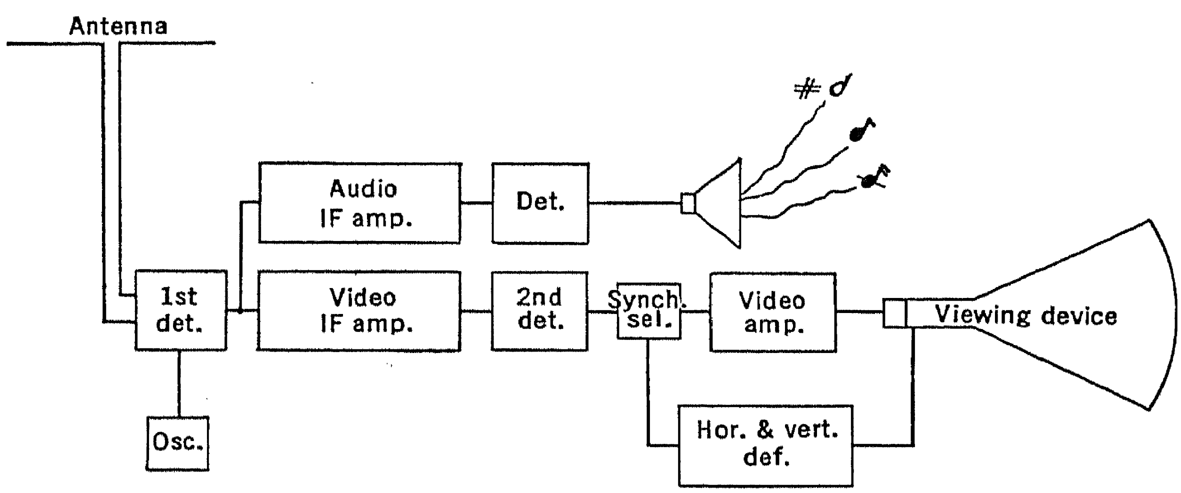


FIG. 7.10.—Block Diagram of a Superheterodyne Television Receiver.

toring amplifier. A superheterodyne receiver is illustrated by the block diagram given in Fig. 7.10.

7.6. Receiver Video Amplifier. The level of the video signal leaving the final detector is, in general, not sufficient to actuate the viewing device. In consequence it must be amplified by several stages of video amplification before it can be used. These stages are similar to those between the pickup equipment and the transmitter, being resistance-coupled and compensated to increase the high-frequency response. Because of the nature of the coupling, the amplifier will pass only the a-c components of the video signal. Therefore it is usually necessary to provide some arrangement to convert the signal to a d-c signal, in order to establish the average illumination of the picture being transmitted. By "d-c" signal is meant one for which the voltage relative to some fixed reference voltage corresponds to the absolute brightness of the point

under the exploring spot at the transmitter. The diagrammatic oscillograms of Fig. 7.11 illustrate the distinction between the two types of signal. The d-c level of the video signal can be reinserted by making use of the pedestal or of the synchronizing impulse, because these bear a definite relation to either maximum dark or light of the picture. A diode rectifier or its equivalent in connection with the final stage of the amplifier, as will be described in Chapter 17, is the usual means for reinserting the d-c component of the signal.

7.7. The Viewing Device. The viewing device converts the video signal into the final visible reproduction which is the ultimate objective of the system.

Two operations are necessary to perform the reconstructing of the image. The first consists of modulating a small but intense spot of light with the video signal; the second, of causing this spot to move in such a way that it sweeps out the scanning pattern on the viewing screen.

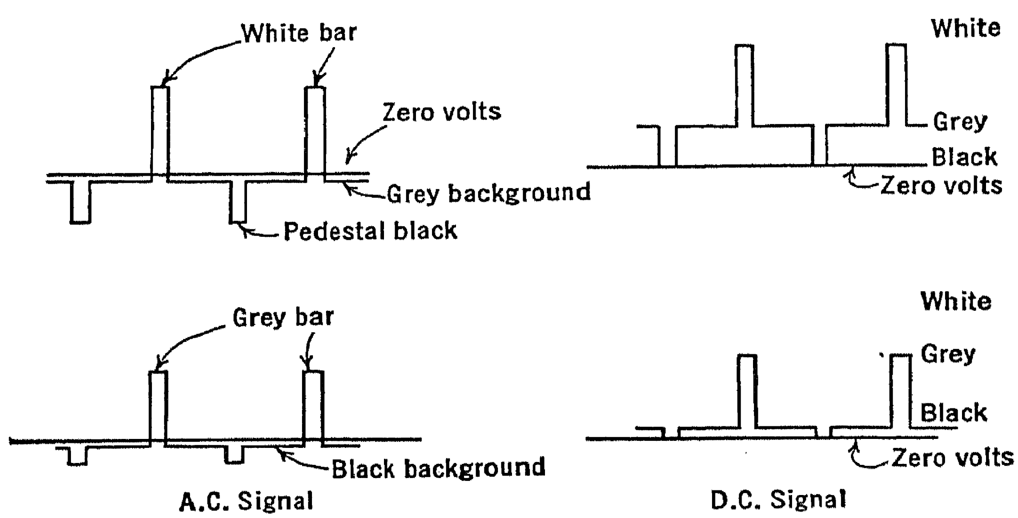


FIG. 7.11.—The Video Signal with and without the D.C. Component.

Broadly, there are two classes of viewing devices, namely, electronic and mechanical. Of these, the former is the most widely used, particularly in home receivers. Very interesting results have been obtained with both systems as a means of producing large pictures for public viewing.

A survey of electronic and mechanical viewing devices will be presented in Chapter 9. Further description of these devices and their operation will be deferred until then.

7.8. Synchronization. As was pointed out in the preceding chapter, the geometric fidelity of the reproduction depends upon the exact correspondence in position of the scanning spot at the transmitter and receiver. This correspondence requires that the periodicity and phasing of the horizontal and vertical scanning motions at the two termini be alike.

It has not been found practicable to attempt to maintain this condition with independent control of the scanning pattern at the pickup and viewing device. Instead, suitable signals are transmitted which indicate the beginning of each frame (or field repetition in the case of interlaced scanning) and the beginning of each line. These synchronizing signals are part of the complete video signal, and occur during time not utilized by the picture signal itself, that is, during the interval in which the scanning spot is returning to its original position, after completing a line or field traversal. In terms of the frequency spectrum of the picture signal discussed in Chapter 6, the frequency components of the synchronizing signal are multiples of line and frame frequency, their phase and amplitude being such that they have no effect on the picture except around the edges where they form a narrow border which cannot be used for the image, and which, in cathode-ray terminal tubes, has no real existence. Instead of this multiplicity of frequencies it might seem possible to use, for purposes of synchronization, a single frequency, falling, for example, in one of the vacant regions of the picture spectrum. However, while no picture component of this frequency appears in the spectrum, the presence of the signal would be reproduced in the form of a spurious moving shading.

In a simple Nipkow disk mechanical system, the rotation of a single element only need be governed, and for this one kind of impulse alone is necessary. The usual procedure is to drive the disk with two motors; one a variable-speed motor which supplies most of the power and is adjusted to run at nearly the required speed, the other a synchronous motor driven by the synchronizing impulse itself, which serves to lock the disk into the correct speed and phase.

More often two types of synchronizing impulses are needed, one to control the horizontal spot displacement, the other the vertical motion. Mechanical film scanning requires synchronization of the vertical scanning produced by the motion of the film, and of the disk or drum giving line scanning. Similarly, where two mirror drums are used for picture reproduction, one for each direction of scanning, both frame and line synchronization are necessary.

Scanning in the case of electronic terminal tubes is produced by deflecting an electron stream periodically in two mutually perpendicular directions by means of suitably varying magnetic or electrostatic fields. The current or voltage producing these fields is supplied from two deflection generators, one operating at line frequency, the other at frame (or field) frequency. Each generator is controlled by its own synchronizing impulse; therefore, the complete signal must include two types which can be distinguished from one another by some form of selector circuit.

It is general practice to make the generator producing the synchronizing impulses at the transmitter the fundamental timing unit of the entire system. Thus the synchronizing signal governs not only the scanning pattern at the reproducer, but also that at the pickup as well. When possible, the synchronizing signal generator is itself tied in with the public electric power supply servicing the area over which the transmitter is being used, in order to minimize the effect of interference between the picture and the 60-cycle power supply.

The exact shape of the impulses for horizontal and vertical synchronizing depends upon how the signal is applied at the deflection generator, and upon the circuits which separate the two components. This phase of the problem will be considered in a later chapter.

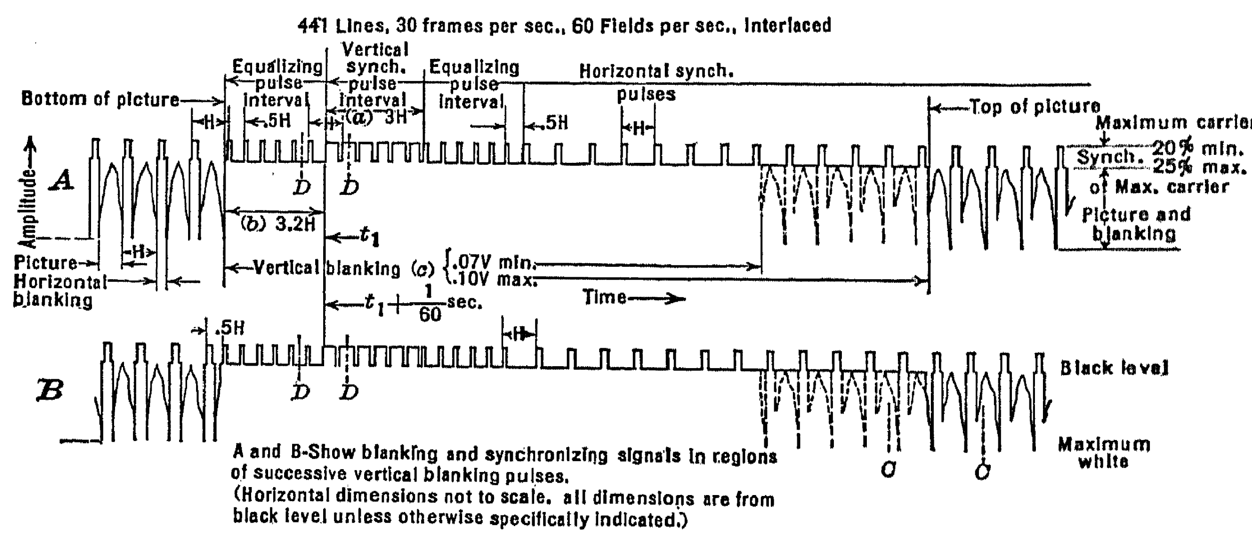


FIG. 7.12.—RMA Standard Television Synchronizing Signal.

Experience has prompted the almost universal adoption of synchronizing impulses which are "blacker than black." That is, the blanking level corresponds to black in the video signal, and the impulses extend below this level in the direction of black.

Some of the conditions which must be satisfied by the synchronizing signal are:

1. The wave shape of the pulse must be such as to permit accurate control of the scanning pattern.
2. The pulse must not interfere with the picture.
3. The signal must be of such shape and amplitude that synchronization is not easily interrupted by interference.
4. The line and frame pulses must be easily separable.
5. Where possible, the signals should aid in some other picture operation, such as extinguishing the beam in the cathode-ray viewing tube or providing automatic gain control.

The form of the synchronizing impulse which has been adopted as the present standard in this country by the R.M.A. is shown in Fig. 7.12. How this signal meets the listed requirements will be discussed in Chapter 15, devoted to the problems of scanning and synchronization.

REFERENCES

1. F. Schroeter (editor), "Fernsehen," Springer, Berlin, 1937.
2. F. E. Terman, "Radio Engineering," McGraw-Hill, New York, 1937.
3. R. D. Kell, "Description of Experimental Television Transmitting Apparatus," *Proc. I. R. E.*, Vol. 21, pp. 1674–1691.
4. G. L. Beers, "Description of Experimental Television Receiver," *Proc. I. R. E.*, Vol. 21, pp. 1692–1706.
5. "An Experimental Television System," *Proc. I. R. E.*, Vol. 22, November, 1934. Part I by E. W. Engstrom, pp. 1241–1245; Part II by R. D. Kell, A. V. Bedford, and M. A. Trainer, pp. 1246–1265; Part III by R. S. Holmes, W. L. Carlson, and W. A. Tolson, pp. 1266–1285.
6. W. J. Poch and D. W. Epstein, "Partial Suppression of One Side-Band in Television Reception," *Proc. I.R.E.*, Vol. 25, pp. 15–31, January, 1937; also *R.C.A. Rev.*, Vol. 1, pp. 19–34, January, 1937.
7. J E. Smith, B. Trevor, and P. S. Carter, "Selective Sideband and Double Sideband Transmission of Telegraph and Facsimile," *R.C.A. Rev.*, Vol. 3, pp. 213–238, October, 1938.

CHAPTER 8

VIDEO PICKUP DEVICES

8.1. Mechanical Systems: The Nipkow Disk. The earliest means for obtaining scanning suitable for television transmission was the Nipkow scanning disk. In its simplest form it is a perforated disk, having apertures spaced at equal angular intervals on a spiral. A Nipkow disk having a spiral of one turn is shown in Fig. 8.1.

The basic arrangement of a Nipkow disk in a direct pickup device for a mechanical television system is illustrated in Fig. 8.2. The scene being transmitted is imaged on the disk by means of the lens L. Behind the disk is a light-sensitive element such as a photosensitive secondary-emission multiplier, or a phototube, which generates a current proportional to the light passing through the apertures. The image area occupies only a small portion of the disk's surface, its width being equal to the chord separating the apertures. The top and bottom of the image coincide with the outer and inner apertures. One and only one opening at any given instant permits light to reach the light-sensitive element. As the disk turns, successive apertures sweep across the image area, scanning it with a pattern of nearly straight parallel lines.

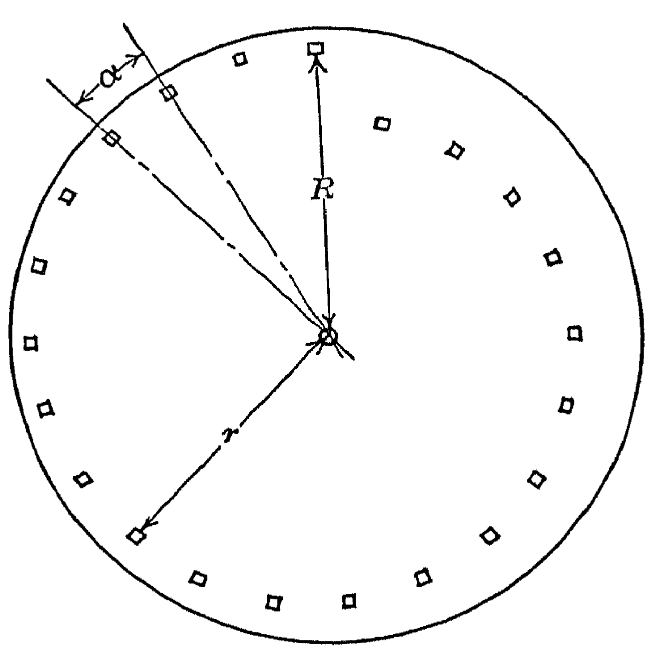


FIG. 8.1.—Nipkow Disk.

With a spiral of one revolution as shown, the disk must rotate at frame frequency N. Furthermore the angular separation between the holes must be $\alpha = 360/n$ degrees, where n is the number of lines making up the scanning pattern. The spacing between these perforations is therefore $2\pi r/n$, r being the radial distance from the center of the disk to the point in question. This spacing is the width of the image on the disk. Assuming an aspect ratio of $1/b$ for this image, its height must be $2\pi br/n$. The perforations for maximum efficiency are square, and so positioned that the edges of consecutive lines just touch. Obviously the height of each square

opening is therefore $2\pi br/n^2$. It should now be apparent that r can be taken as equal to the radius R of the disk, since over the entire pattern all the perforations lie within R and $R\left(1 - \frac{2\pi b}{n}\right)$, and n for high definition is large.

The usefulness of this pickup arrangement depends upon the video signal which can be obtained with a given amount of light. Referring again to Fig. 8.2, the light reaching the photosensitive element S, placed behind the disk, and generating the video signal, can be estimated as follows. On a very bright midsummer day the illumination in an outdoor

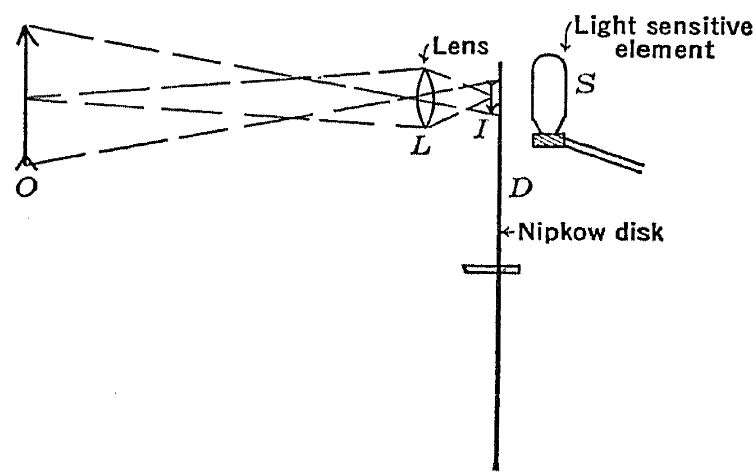


FIG. 8.2.—Direct Pickup Using a Nipkow Disk.

scene may reach as high as 10,000 foot-candles, although 3000 foot-candles might be taken as an average for summer daytime illumination. If a reflection coefficient of ½ is assumed, the brightness B of the scene will be:

$$B = \frac{10{,}000}{2\pi} \cong 1500 \text{ candles/sq. ft.}$$

The incident illumination of the image projected on the disk depends upon the photographic aperture or f number A of lens L. For purposes of computation a value of $A = 2$ for the aperture is assumed. The illumination under these conditions is:

$$I = \frac{\pi B}{4A^2} = \frac{10{,}000}{2\pi}\,\frac{\pi}{4 \cdot 2^2} \cong 300 \text{ lumens/sq. ft.}$$

It should be noted that absorption in the lens is not considered, and the field angle is assumed to be small. Practical considerations limit the maximum diameter of the scanning disk, for it must be maintained accurately in synchronism at a very high angular velocity. A circumference

of 12 feet, though large, is not beyond physical realization. The light flux passing through an aperture of a disk of this size will be:

$$F = I\left(\frac{2\pi br}{n^2}\right)^2 = 4 \times 10^4 \cdot \frac{b^2}{n^4} \text{ lumen.}$$

If, now, a picture of the quality under discussion is being transmitted, with 441 lines making up the scanning pattern, the amount of light reaching the phototube will be about 6×10^{-7} lumen. Even if the sensitivity of the light-sensitive element is 50 microamperes per lumen, the signal current will be only 3×10^{-11} ampere. Whether or not this amount of signal is sufficient to give an acceptable picture depends upon the noise fluctuation that accompanies the signal. The absolute minimum of noise is determined by the statistical fluctuation of the photocurrent itself. This lower limit might be approximated if the light-sensitive element were a good secondary-emission multiplier. This limiting noise, over the required 4-megacycle video band, is:

$$\overline{i_n{}^2} = 2\,eIf = 32 \times 10^{-20} \times 3 \times 10^{-11} \times 4 \times 10^6$$

or:

$$\overline{i_n}(\text{rms}) \cong 6 \times 10^{-12} \text{ ampere.}$$

In other words, the signal-to-noise ratio, even under the ideal illumination assumed, is only about 5. This ratio is far too low to give what can be considered a picture of high entertainment value.

The light reaching the light-sensitive element can be increased to some extent by means of a more complicated arrangement, as, for example, a disk having a spiral of two or more turns, or one in which lenses replace the apertures. However, this increase is not sufficient to make the system practical for direct pickup.

Although these calculations indicate that the scanning disk is not suited for this service, it finds its place in other types of pickup.

8.2. Flying Spot Scanning. For studio pickup of a very small area, e.g., the close-up of a head, a Nipkow disk used in a somewhat different way has been found fairly successful. The pickup equipment is arranged as shown schematically in Fig. 8.3. Instead of an image of the scene being projected on the disk, the disk is illuminated with a high-intensity arc, or other powerful light source, so that the light through it forms a narrow moving beam. This beam is focused on the object being televised, to produce a spot of light which scans its surface. The light reflected back from the object is received in a battery of phototubes surrounding it. These tubes produce the video signal. The advantage of imaging the fly-

ing light spot on the scene and picking up the reflected light from all points continuously is twofold. The surrounding phototubes can be made to subtend a very large solid angle about the illuminated point, thus making use of a fairly large fraction of the reflected light. Also, the average illumination needed on the scene is only $1/n^2$ that which would be

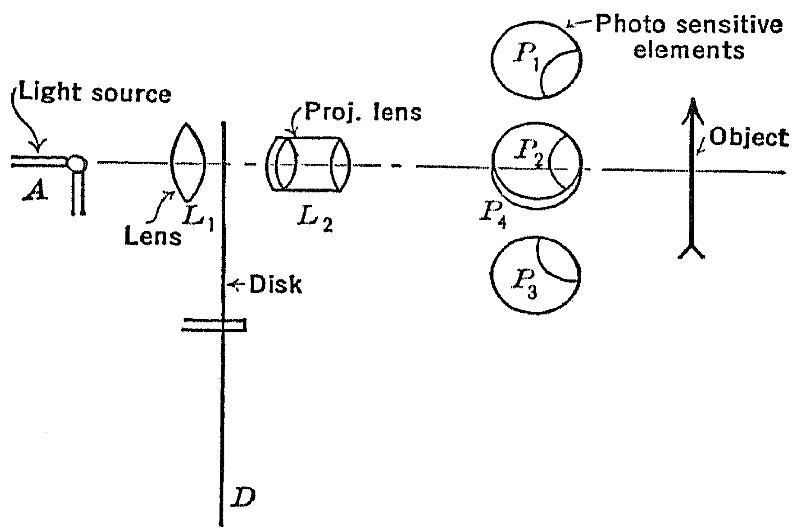


Fig. 8.3.—Flying Spot Scanner.

needed to obtain an equal video signal by the first-mentioned system, even if the light-gathering power of the objective imaging the scene on the disk is assumed to be equivalent to that of the phototubes. In order to maintain the large light-gathering power of the phototubes it is necessary to restrict the field to be reproduced to a relatively small area in the immediate vicinity of the light-sensitive elements.

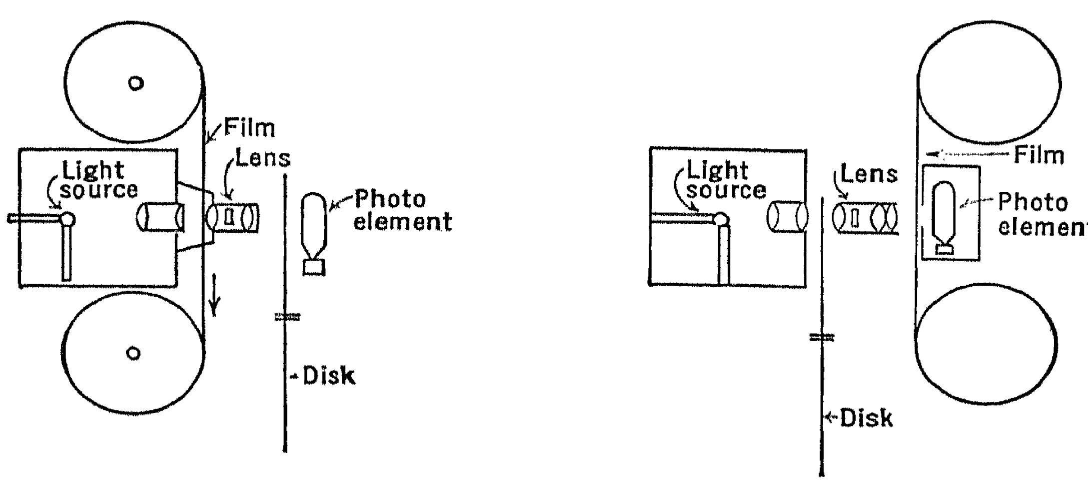


Fig. 8.4.—Two Forms of Simple Disk Film Scanners.

8.3. Film Scanning. In a very similar way the Nipkow disk can be used to pick up the image from a moving-picture film. Two arrangements for this type of pickup are shown in Fig. 8.4. The image from the film is

projected onto the scanning disk in the first. In the other the light source is imaged onto the disk, and the opposite side of the disk is in turn imaged onto the film, so that the apertures produce a small moving spot of light which sweeps over the picture. In each case, the light passing through the disk and film is picked up and converted into the video signal by a light-sensitive element. The choice between these two arrangements rests on practical considerations of mechanical construction, rather than any inherent superiority in optics. For the direct pickup of actual objects, vertical displacement of the scanning spot is obtained through arranging the series of apertures in a spiral. This is not necessary for film pickup, as the motion of the film itself can be utilized. The perforations in the disk are arranged in a circle, and the scanning spot moves across a single line,

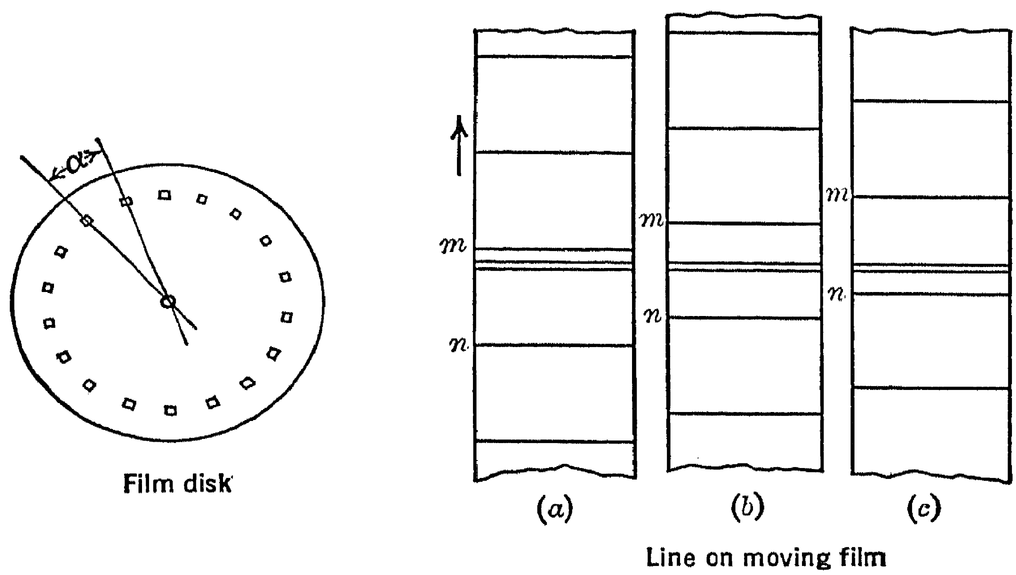


FIG. 8.5.—Principle of Film Scanning.

rather than over a pattern of parallel lines. The film is moved with constant speed and at such a rate that a single frame passes the position of the scanned line in one picture period. Therefore, as illustrated in Fig. 8.5, relative to the film the scanning spot sweeps out an ordinary pattern of parallel lines.

So far no mention has been made of interlaced scanning. Many expedients are possible. For film transmission, odd-line interlacing is obtained if the disk contains the same number of apertures as the complete pattern has lines, and the film passes it at the rate of two frames per revolution. A double spiral of holes will give interlacing for studio pickup.

To return to film scanning, the disk can be designed so as to make more than one revolution per picture period. If α is the angular separa-

tion between apertures, and n' the number, the relation between the number of revolutions per second N' and the frame frequency is:

$$\alpha n N = 2\pi N',$$
$$\alpha n' = 2\pi,$$
$$\frac{n}{n'} N = N'(\text{rps}).$$

When n' is less than n, a higher rate of revolution is required; however, the diameter of the disk can be reduced, with a consequent reduction in the net mechanical stresses.

8.4. Modification of Scanning Disk. As an alternative to the disk described, a drum perforated with a helical array of apertures can be used to obtain the scanning pattern required for the direct pickup from a physical object. Similarly, a single row of perforations around such a drum will serve as a film scanner

A modification of the simple disk is shown in Fig. 8.6. Instead of the tiny apertures heretofore described, the holes are fairly large and each is fitted with a lens. Fig. 8.6 shows a lens disk arranged for film pickup. Each lens in the disk is a micro-objective of very short focal length and large numerical aperture. These objectives form very much reduced images of the relatively large aperture on the film. The tiny bright image is the exploring element which sweeps out the scanning pattern. Again a phototube intercepts the light passing through the film and generates the video signal. Of course, the lens disk has its counterpart in the lens drum. A lens drum used for practical television transmission by Telefunken is shown photographed in Fig. 8.7. This drum has a light gain of the order of 30 times over a simple perforated drum of equal size. The drum is equipped with a single row of holes for film transmission and a helix of two revolutions for direct pickup. The complete scanner incorporating this drum is illustrated in Fig. 8.8.

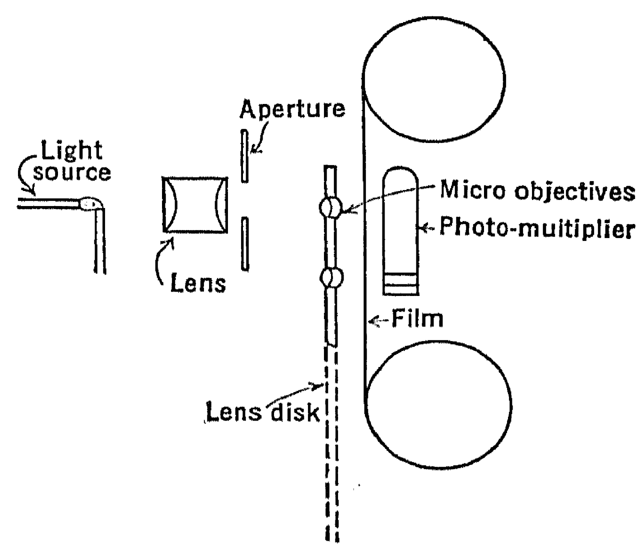


FIG. 8.6.—Film Scanner with Lens Disk.

A very successful film transmitter, built and used by the Fernseh Company, is shown in Fig. 8.9. A 375-line picture with a 25-cycle frame frequency is obtained from it. In this machine an image of the continu-

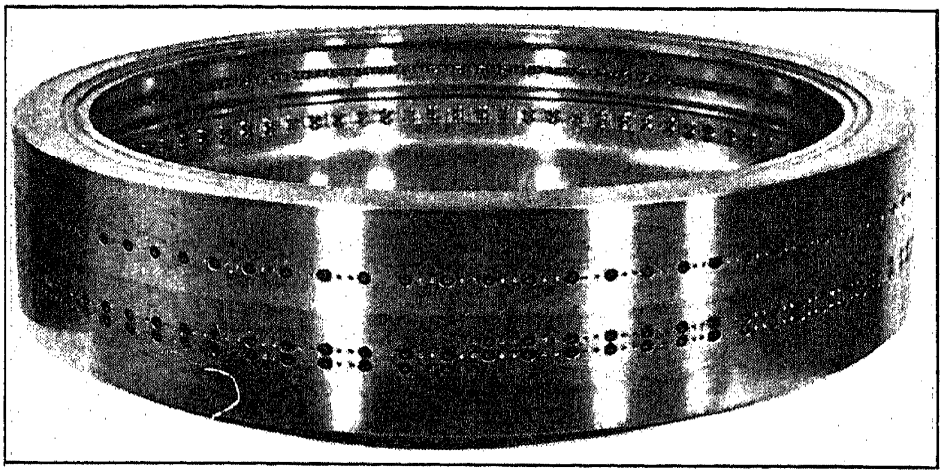


Fig. 8.7.—Lens Drum (Telefunken). (Courtesy of J. Springer.)

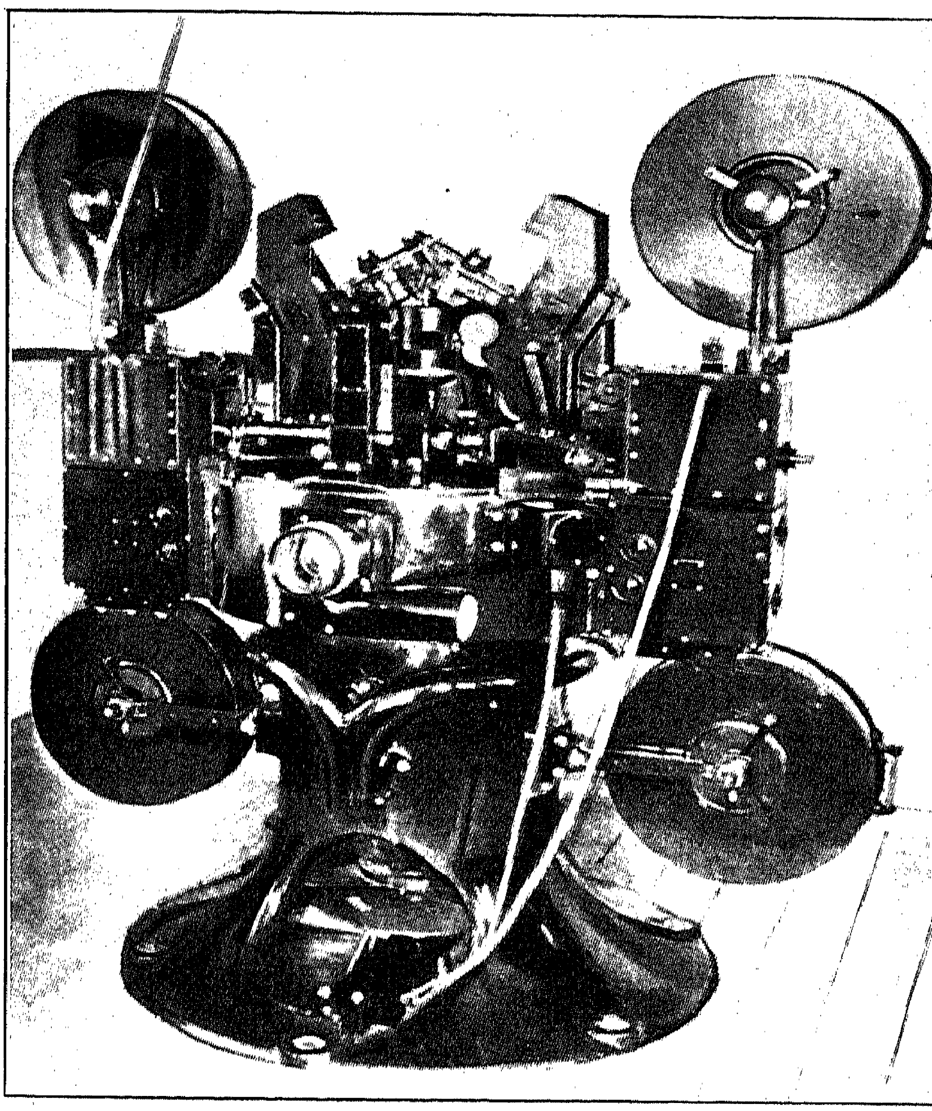


Fig. 8.8.—Complete Film and Direct Pickup Scanner (Telefunken). (Courtesy of J. Springer.)

ously moving film is projected onto an aperture disk, and the light passing through the latter is picked up by a photoelectric secondary-emission multiplier. The disk itself is very carefully balanced, and turns in a partially evacuated housing in order to reduce the power required to drive it and thus facilitate synchronization.

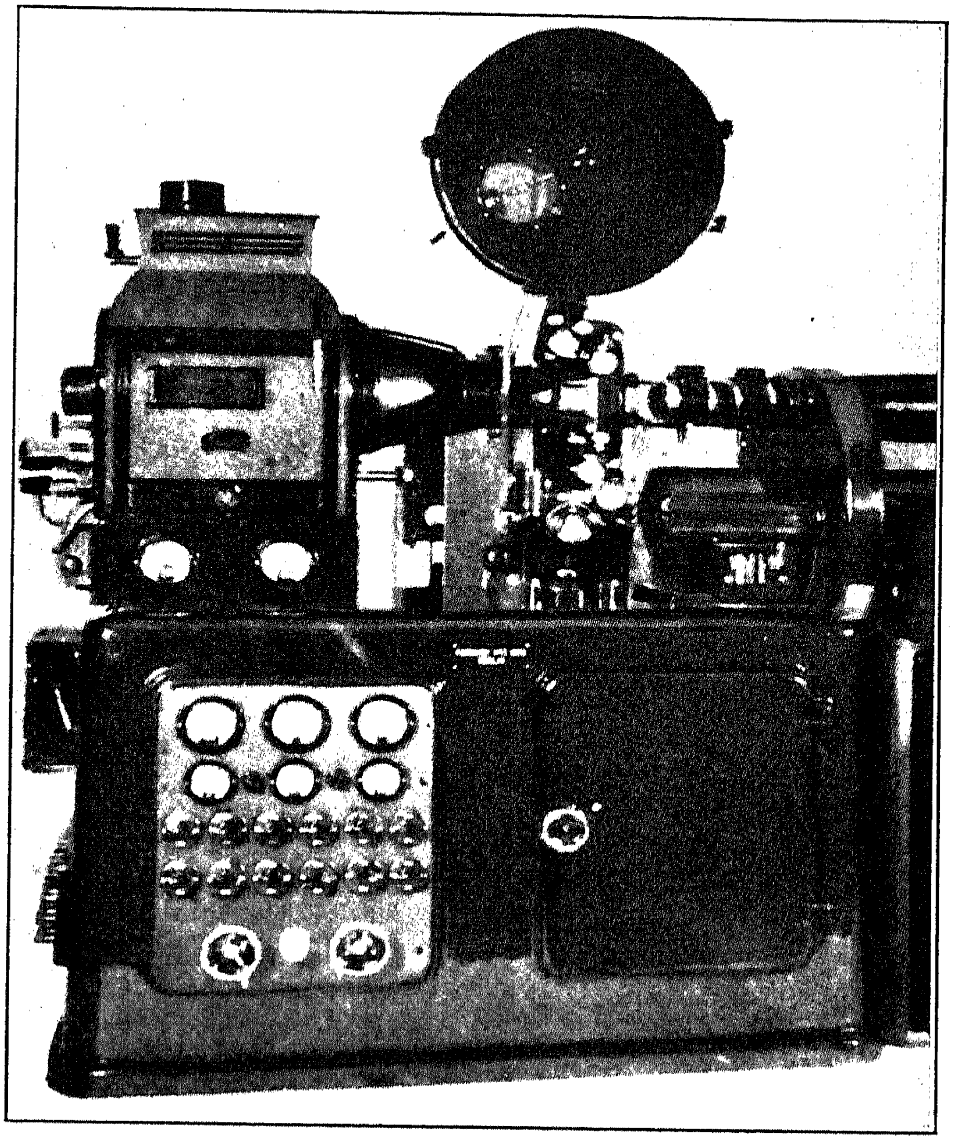


Fig. 8.9.—Lens Disk Film Scanner (Fernseh). (Courtesy of J. Springer.)

The video-telephone system now operating in Germany between Berlin, Leipzig, and Munich makes use of a lens-disk flying spot scanner giving a 180-line pattern. Large photoelectric multipliers arranged around the window through which the moving spot is projected transform the reflected light into the video signal. This scanner is shown in Fig. 8.10.

Other forms of mechanical scanning have been used experimentally with varying degrees of success. The mirror drum, illustrated in Fig. 8.11, exemplifies one form. A modification of this drum has curved mirrors

instead of flat. Prisms take the place of mirrors in another, very similar, arrangement. These scanners, together with others not mentioned, offer some promise as a means of effecting mechanical television pickup, but

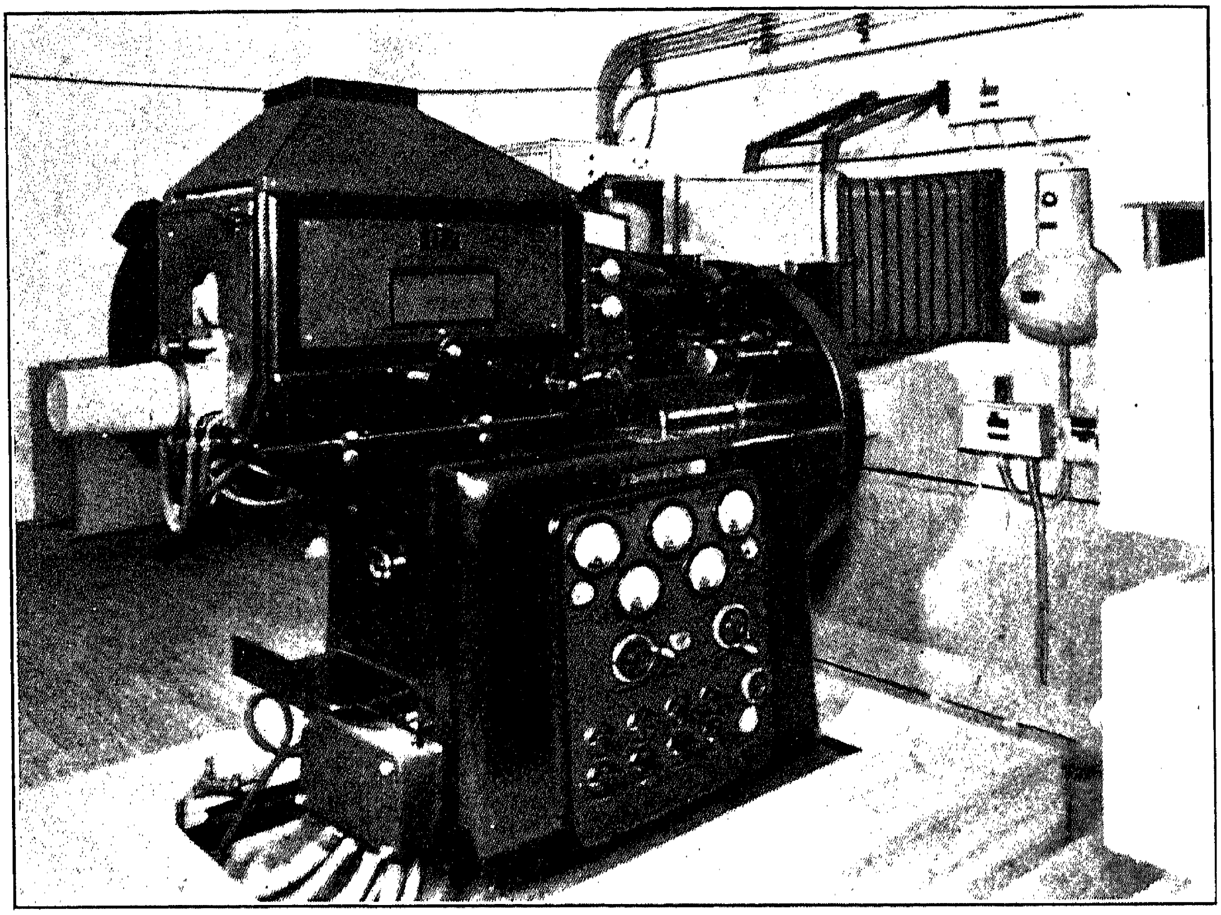


FIG. 8.10.—Flying-Spot Scanner for a Video-Telephone (Fernseh). (Courtesy of J. Springer.)

they have not been developed to the same extent as have the disk and drum scanners described in the earlier pages of this chapter.

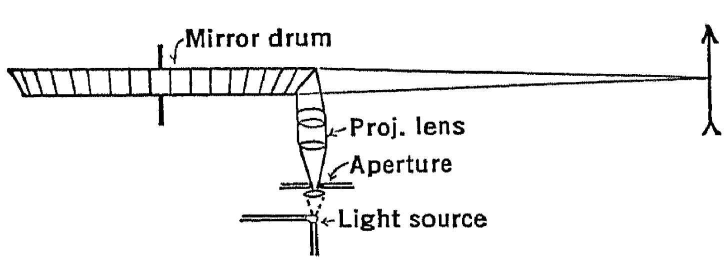


FIG. 8.11.—Schematic Diagram of a Mirror Drum Scanner.

8.5. Intermediate Film Pickup. Although, as the foregoing estimate indicates, the mechanical scanner, because of its intrinsic lack of sensitivity, is excluded from the field of direct pickup, this is not so serious a limitation as it appears to be at first glance. An intermediate moving-

picture film provides the means for overcoming the obstacle. The scene to be televised is photographed by means of a moving-picture camera, rapid development of this film then follows, and finally a mechanical film scanner converts the registered image into the video signal. A diagrammatic view of this type of transmitter, including camera, processing equipment, and scanner, is shown in Fig. 8.12.

One disadvantage of this system is the delay between the actual taking of the picture and converting it into video signal, due to the time required to process the film. Special methods of developing and fixing the film, however, have reduced this time to less than a minute. Of course, it is unnecessary to dry the film before it is used in the scanner. Also, a negative is as satisfactory as a positive for transmission because the reproduced picture can be changed from negative to positive merely by

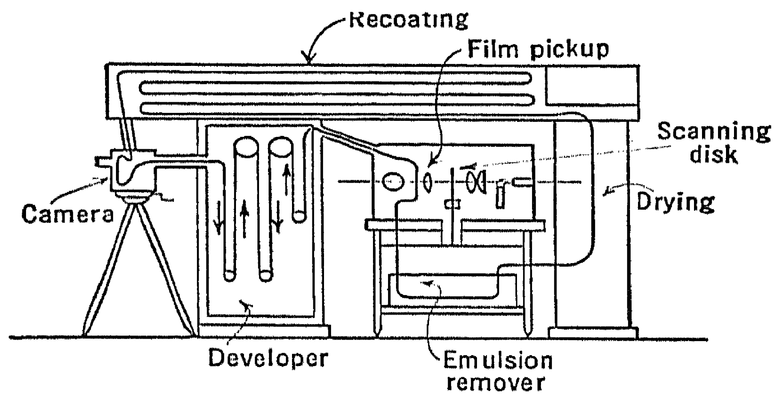


FIG. 8.12.—Continuous Intermediate Film Scanner.

changing the polarity of the signal. (Note: This inversion takes place with every stage of amplification so that an amplifier of an even number of stages gives a picture having the same sense as that at the input, while an odd number of stages inverts light and dark.) Therefore, the delay that would be incurred by printing or reversing the film is avoided. The most serious practical objections to the system are the complicated and bulky equipment required and the cost of operation.

8.6. Electronic Pickup Systems: Electronic Scanning. The first published suggestion of a cathode-ray pickup arrangement appears to have been made by A. A. Campbell Swinton in a letter to *Nature* as early as June, 1908. As a historic landmark this letter is worthy of quoting verbatim:

DISTANT ELECTRIC VISION

Referring to Mr. Shelford Bidwell's illuminating communication on this subject published in NATURE of June 4, may I point out that though, as stated by Mr. Bidwell, it is wildly impracticable to effect even 160,000 synchronised operations per second by ordinary mechanical means, this part of the problem of obtaining distant electric

vision can probably be solved by the employment of two beams of kathode rays (one at the transmitting and one at the receiving station) synchronously deflected by the varying fields of two electromagnets placed at right angles to one another and energised by two alternating electric currents of widely different frequencies, so that the moving extremities of the two beams are caused to sweep synchronously over the whole of the required surfaces within the one-tenth of a second necessary to take advantage of visual persistence.

Indeed, so far as the receiving apparatus is concerned, the moving kathode beam has only to be arranged to impinge on a sufficiently sensitive fluorescent screen, and given suitable variations in its intensity, to obtain the desired result.

The real difficulties lie in devising an efficient transmitter which, under the influence of light and shade, shall sufficiently vary the transmitted electric current so as to produce the necessary alterations in the intensity of the kathode beam of the receiver, and further in making the transmitter sufficiently rapid in its action to respond to the 160,000 variations per second that are necessary as a minimum.

Possibly no photoelectric phenomenon at present known will provide what is required in this respect, but should something suitable be discovered, distant electric vision will, I think, come within the region of possibility.

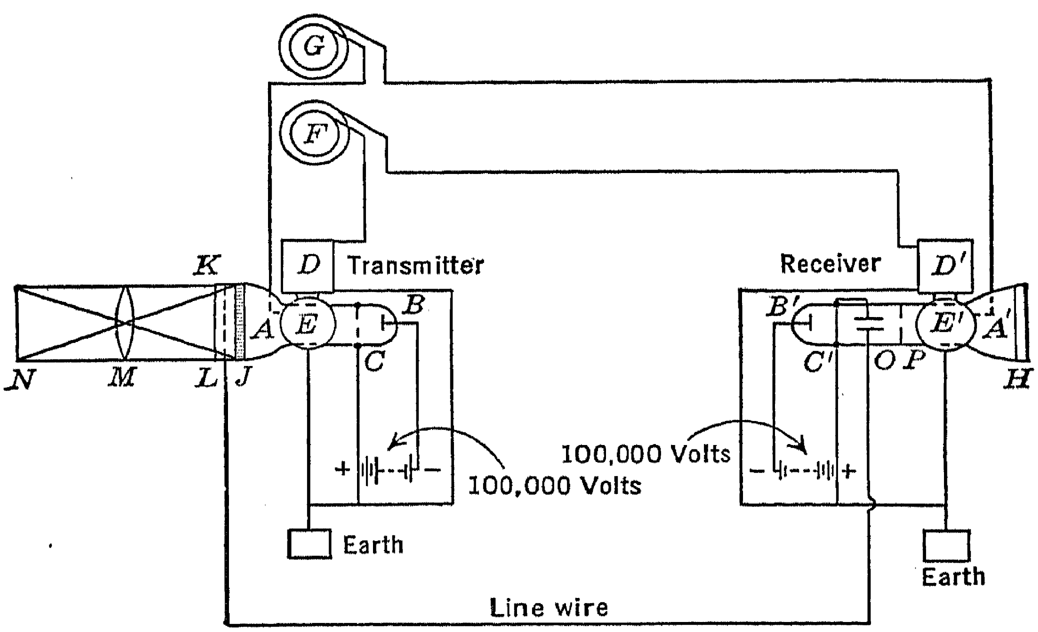


FIG. 8.13.—Electronic Television System Proposed by Campbell Swinton.

Although this letter did not describe in detail how such an electronic scanner could be made to operate to convert the light image into the desired video signal, a later paper by the same author, presented before the Roentgen Society in 1911, and repeated in part in *Wireless World*, April, 1924, discussed in some detail a pickup tube embodying cathode-ray scanning. The tube and circuit are diagrammed in Fig. 8.13. The transmitting tube is a Crookes tube with the cathode *B* connected to the

negative end of a 100,000-volt d-c power source, while the anode is connected to the positive end. The anode is perforated with a fine aperture through which passes a narrow pencil of cathode rays. Continuing with the description as given by A. A. Campbell Swinton in *Wireless World and Radio Review*:*

> . . . the cathode rays fall on a screen J, the whole surface of which they search out every tenth of a second under the influence of the magnets D and E. Further, it is to be remarked that as the two magnets D and D', and the two magnets E and E' are energised by the same currents, the movements of the two beams of cathode rays will be exactly synchronous and the cathode rays will always fall on the two screens H and J on each corresponding spot simultaneously.
>
> In the transmitter, the screen J, which is gas-tight, is formed of a number of small metallic cubes insulated from one another but presenting a clean metallic surface to the cathode rays on the one side, and to a suitable gas or vapour, say sodium vapour, on the other. The metallic cubes which compose J are made of some metal, such as rubidium, which is strongly active photoelectrically, in readily discharging negative electricity under the influence of light, while the receptacle K is filled with a gas or vapour, such as sodium vapour, which conducts negative electricity more readily under the influence of light than in the dark.
>
> Parallel to the screen J is another screen of metallic gauze L, and the image to be transmitted of the object N is projected by the lens M through the gauze screen L on to the screen J, through the vapour contained in K. The gauze screen L of the transmitter is connected through the line wire to a metallic plate O in the receiver, past which the cathode rays have to pass. There is, further, a diaphragm P fitted with an aperture in such a position as, having regard to the inclined position of B', to cut off the cathode rays coming from the latter, and prevent them from reaching the screen H, unless they are slightly repelled from the plate O, when they are able to pass through the aperture.
>
> The whole apparatus is designed to function as follows: Assume a uniform beam of cathode rays to be passing in the Crookes tubes A and A', and the magnets D and E and D' and E' to be energised with alternating current, as mentioned. Assume, further, that the image that is desired to be transmitted is strongly projected by the lens M through the gauze screen L on to the screen J. Then, as the cathode rays in A oscillate and search out the surface of J they will impart a negative charge in turn to all of the metallic cubes of which J is composed. In the case of cubes on which no light is projected, nothing further will happen, the charge dissipating itself in the tube; but in the case of such of those cubes as are brightly illuminated by the projected image, the negative charge imparted to them by the cathode rays will pass away through the ionised gas along the line of the illuminating beam of light until it reaches the screen L, whence the

* See Campbell Swinton, reference 7.

charge will travel by means of the line wire to the plate O of the receiver. This plate will thereby be charged; will slightly repel the cathode rays in the receiver; will enable these rays to pass through the diaphragm P, and, impinging on the fluorescent screen H, will make a spot of light. This will occur in the case of each metallic cube of the screen J, which is illuminated, while each bright spot on the screen H will have relatively exactly the same position as that of the illuminated cube of J. Consequently, as the cathode ray beam in the transmitter passes over in turn each of the metallic cubes of the screen J, it will indicate by a corresponding bright spot on H whether the cube in J is or is not illuminated, with the result that H, within one-tenth of a second, will be covered with a number of luminous spots exactly corresponding to the luminous image thrown on J by the lens M, to the extent that this image can be reconstructed in a mosaic fashion. By making the beams of cathode rays very thin, by employing a very large number of very small metallic cubes in the screen J, and by employing a very high rate of alternation in the dynamo G, it is obvious that the luminous spots on H, of which the image is constituted, can be made very small and numerous, with the result that the more these conditions are observed, the more distinct and accurate will be the received image.

Furthermore, it is obvious that, by employing for the fluorescent material on the screen H something that has some degree of persistency in its fluorescence, it will be possible to reduce the rate at which the synchronised motions and impulses need take place, though this will only be attained at the expense of being able to follow rapid movements in the image that is being transmitted.

It is further to be noted that as each of the metallic cubes in the screen J acts as an independent photoelectric cell, and is only called upon to act once in a tenth of a second, the arrangement has obvious advantages over other arrangements that have been suggested, in which a single photoelectric cell is called upon to produce the many thousands of separate impulses that are required to be transmitted through the line wire per second, a condition which no known form of photoelectric cell will admit of.

The basic similarity between this type of tube and the mechanical scanners is evident from the description of its mode of operation. Instead of there being a single photocell upon which light is projected from the various picture elements in succession, there are many photocells which are in operation only during the time the scanning beam, which functions as a commutator, is actually on them. The photocurrent which makes up the video signal is collected on an anode which is common to all the cells, but since only one cell is active at any one instant, the current collected corresponds to that from the successive elements as the beam sweeps over them. The limiting sensitivity can be calculated in the same way as that for the disk scanner. If A is the area of the array of elements and there are n^2/b elements in all, corresponding to an n-line picture

with an aspect ratio of $1/b$, the area of each element a will be Ab/n^2. Now, assume an average illumination I falling on the picture area, and an effective photosensitivity p. The average current picked up by the collector element L (since one element, and only one element, is always under the beam) is given by:

$$i_s = Ip\frac{bA}{n^2}.$$

Under the conditions previously assumed with a screen area of 20 square inches and a photosensitivity of 20 microamperes per lumen, the signal current is 3.2×10^{-9} ampere. If this current could be collected with

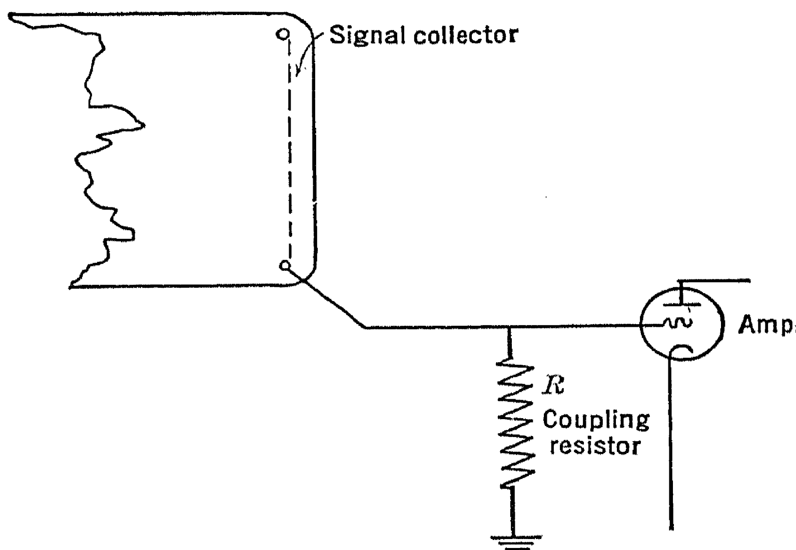


FIG. 8.14.—Simple Means for Coupling a Pickup Tube to the Video Amplifier.

an ideal multiplier, the accompanying noise would be about 6×10^{-11} ampere. The signal-to-noise ratio, therefore, is sufficient to give a picture of the desired quality. The ratio, which decreases with the square root of the illumination, becomes too low for practical purposes when the illumination falls below an average value of 1000 foot-candles. Even at that value, noise would be plainly visible in the picture.

In computing the limiting sensitivity of the mechanical pickup arrangement it was assumed that the photocurrent was amplified by means of a "noiseless" multiplier. Such an arrangement is not always practical. If, instead of the secondary-emission multiplier, a thermionic amplifier of conventional design is employed, the limiting sensitivity is greatly curtailed. This is because the statistical fluctuation introduced by the amplifier and coupling arrangement is much greater than that present in the photocurrent. The signal current output from the pickup device is passed through the coupling resistor, as shown in Fig. 8.14, and the IR drop across it actuates the grid of the first radio amplifier tube. The mag-

nitude of this voltage depends upon the value of the coupling resistor, and this in turn is limited by the capacity to ground of the combined output element of the pickup tube, coupling element, and amplifier grid. Details of this limitation will be taken up in Chapter 14. With the electronic device just described, if a resistor of 20,000 ohms is used, the signal voltage is:

$$e_s = 0.6 \times 10^{-4} \text{ volt,}$$

while the noise voltage, as will be shown in the chapter referred to, is:

$$\overline{e_n}(\text{rms}) = 3.5 \times 10^{-5} \text{ volt.}$$

Under these circumstances the signal-to-noise ratio is below the useful limit. A somewhat more complicated coupling arrangement may increase

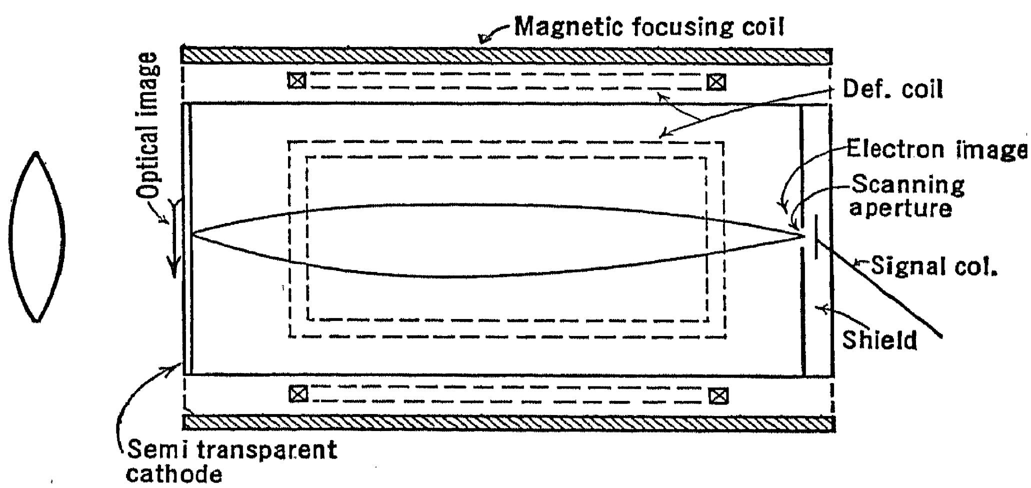


Fig. 8.15.—An Early Farnsworth Dissector Tube.

this ratio to some extent, but not greatly. The idea of replacing mechanical moving parts by an electron beam, however, represents a very important step forward.

8.7. Dissector Tube. Another and quite different type of electronic scanning was developed by Philo T. Farnsworth.* The principle of this pickup device is shown in Fig. 8.15. The optical image is projected onto the photocathode, from which are released electrons whose density is distributed in accordance with the distribution of light intensity. Thus, an electron image is formed. These electrons are accelerated by a potential difference between the photocathode and the shield at the other end of the tube (or suitable metallic coatings on the wall, etc.) and move toward the shield. A magnetic field produced by a long coil focuses the electron image on the shield. This shield is perforated in the center by a tiny aperture behind which is an electrode which collects the electrons passing through the aperture. Two pairs of coils, perpendicular

* See Farnsworth, reference 8.

to the axis of the focusing field and to each other, deflect the image as a whole, either vertically or horizontally. By supplying one pair with a sawtooth current at line frequency, and the other with a similar current wave at frame frequency, the entire image is moved in such a way that

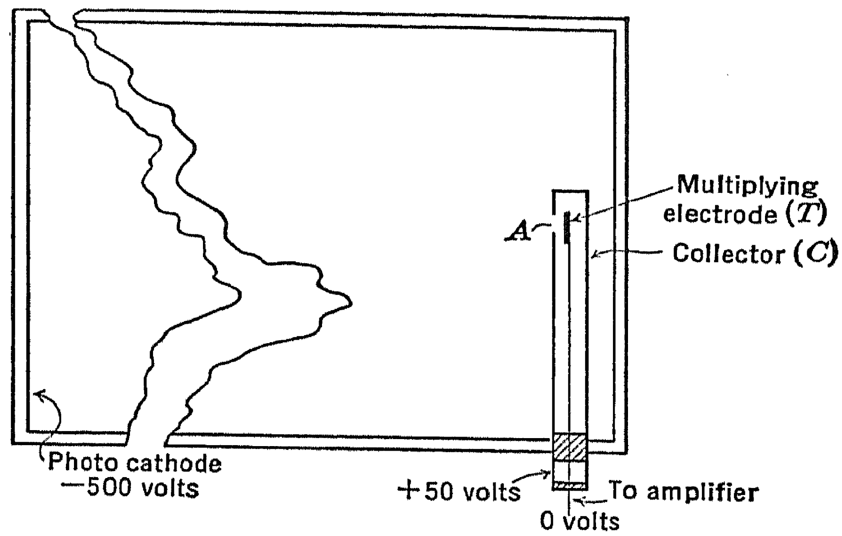


FIG. 8.16.—Simple Multiplier in Farnsworth Dissector Tube.

the aperture sweeps out a scanning pattern relative to the image itself. In other words, moving the image across the scanning aperture is exactly equivalent to moving the aperture over the image. The current to the collecting electrode has an instantaneous value equal to the

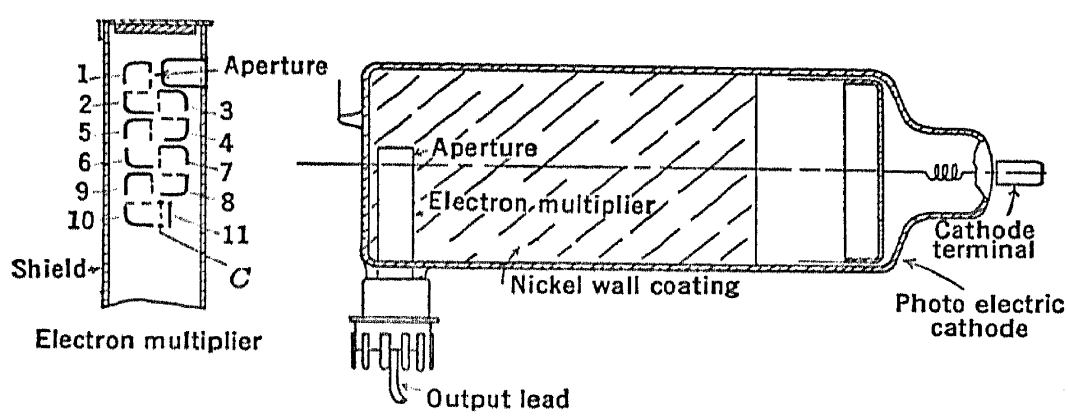


FIG. 8.17.—Diagram of Modern Farnsworth Dissector Tube. (Courtesy Farnsworth Television, Inc.)

photoemission from the picture element resting on the aperture, and therefore varies as the aperture scans the image, to give the video signal. Just as in the previous pickup arrangement proposed by Campbell Swinton, the signal current has an instantaneous value equal to the photocurrent from a single element.

The electron optics involved in this device, which goes under the name

of Dissector Tube, is of no little interest. The problem is one of imaging the emission from an extended photocathode. However, it is not identical with that of the image tube mentioned in Chapter 4, because in the latter all image points must be simultaneously in focus, whereas in the Dissector Tube only those points in the immediate vicinity of the scanning aperture need be accurately sharp. This makes it possible to use suitable components of the deflecting voltage or current to correct for any curvature of the image field when necessary. These components may be supplied either to the magnetic lens or to the accelerating voltage.

The Dissector Tube used with a conventional amplifier and coupling

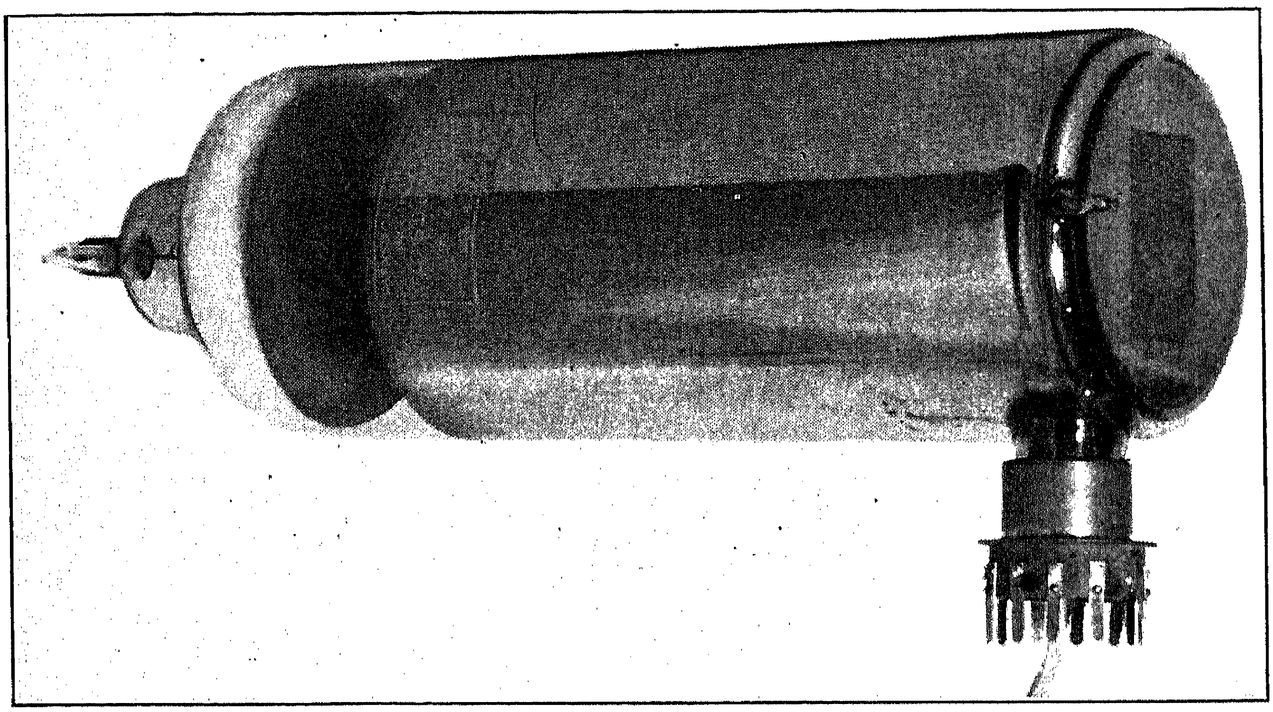


Fig. 8.18.—Photograph of Modern Farnsworth Dissector Tube. (Courtesy Farnsworth Television, Inc.)

system, such as was described for the first-mentioned electronic pickup tube, would yield a signal-to-noise ratio too low to produce a picture of high entertainment value, except under exceptional lighting conditions, or in conjunction with a moving-picture projector. However, this type of tube lends itself readily to the use of a secondary-emission amplifier. As first used in this type of tube the multiplier took the form shown in Fig. 8.16. Instead of the shield shown in Fig. 8.15, a narrow hollow metal tube contains the aperture which scans the image. This permits using an opaque photoemitter instead of the semi-transparent or screen cathode, the light being only slightly obstructed by the metal tube. Inside the tube, directly behind the aperture, is a small secondary emitting element T which is made a few volts negative with respect to the tube itself. Electrons entering the aperture strike this element, producing secondary

electrons which are collected by the walls of the metal tube. The net current resulting from the difference between the secondary-emission current and the entering electron stream is the signal current. With a good emitter for the target T, the current will be eight or more times the incoming photocurrent. Further secondary-emission multiplication was obtained in later tubes by incorporating a minute dynamic secondary-emission multiplier in the tube containing the aperture. The present form of this tube uses a multi-stage static multiplier. A tube of this type is shown in Figs. 8.17 and 8.18. In practice these tubes are found to give an excellent television image when the illumination is adequate. Because of the fairly high incident illumination needed, they are rather better suited for film reproduction than direct pickup, although they are in no way excluded from practical use in the latter field.

Electronic scanning in the viewing tube, as successfully demonstrated by Boris Rosing in 1907 and in the pickup tube as proposed by Campbell Swinton in 1908, were the first two steps leading up to the modern electronic high-definition system. The third important step was the introduction of the storage principle.*

8.8. Storage Principle—The Iconoscope. In the systems described heretofore, the photoemission from each picture element is used only during the very short interval of time it is actually covered by the scanning spot or aperture. In contradistinction, under the storage principle, photoemission continues during the entire time, being accumulated in the form of charge at each image point. The entire accumulated charge is removed once each picture period, from each of the picture elements in sequence, as the scanning spot passes over them. The result is an increase in the effective photocurrent by a factor equal to the number of picture elements.

This principle, together with that of scanning with a cathode-ray beam, forms the basis of the Iconoscope. There are many possible modifications of this tube. Fig. 8.19 illustrates one form whose operation is straightforward and therefore easily explained. This Iconoscope consists of an electron gun G, a mosaic screen S, and two collecting elements A_2 and C, all enclosed in a glass bulb which is thoroughly evacuated. The details of the mosaic are shown in the figure. A conducting mesh screen is coated with a thin dielectric layer, leaving the openings of the mesh free. These interstices are filled with plugs of a photoelectric material. Silver plugs activated with caesium are often used. The mosaic is mounted in the tube in such a position that the photosensitized side of the plugs faces the collector C, and the opposite side faces the gun.

The scanning beam, when it rests on any element, produces secondary

* See Zworykin, reference 9.

electrons, and if the emission is saturated the yield exceeds the primary beam. Since the element is insulated it tends to assume a positive equilibrium potential with respect to the collector A_2 such that the current reaching the element just equals the current leaving. This equilibrium potential is independent of the initial potential of the element.

When a portion of the mosaic is illuminated, elements in that area emit photoelectrons which go to the positive collector C. By virtue of this emission the elements accumulate a positive charge. When the scanning beam in the course of sweeping out the pattern passes over a positively charged element it is returned to its original equilibrium potential. Each illuminated element therefore goes through the cycle of accumu-

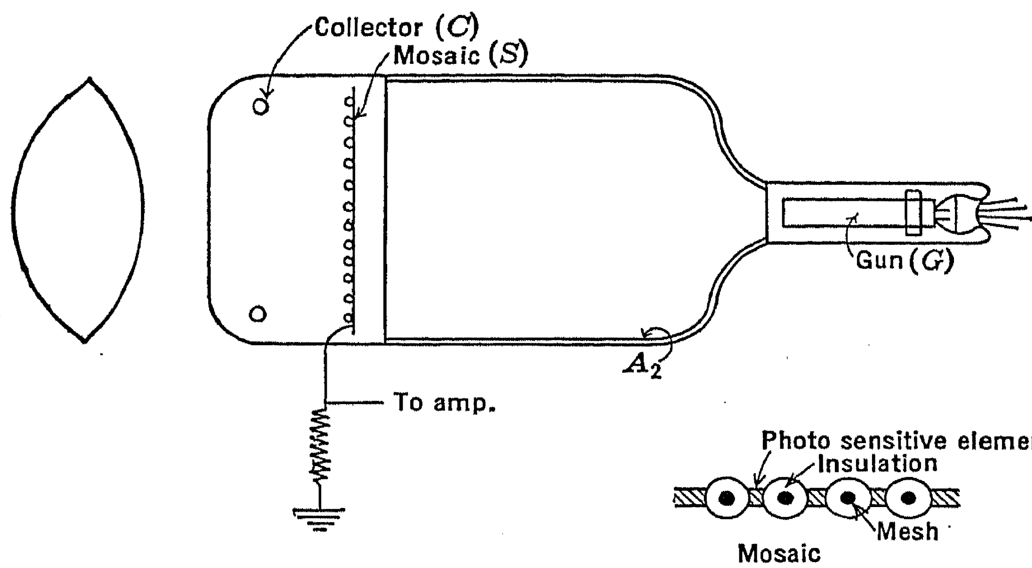


FIG. 8.19.—Diagram of Iconoscope Employing a Two-Sided Mosaic.

lating a positive charge during the entire picture period and releasing it each time the scanning beam sweeps over it. Since each element is coupled by capacity to the metallic screen which forms the foundation of the mosaic, any change in charge on the former induces a corresponding change in the latter. The change in charge produces a current in the lead connected to the coupling impedance and amplifier.

The sensitivity of this system can be estimated in a manner similar to that used to determine the behavior of the preceding systems. If I is the average light intensity of the optical image, A the area of the mosaic, and n^2/b the number of elements, the charge accumulated by an element during one picture period is:

$$q = \frac{IApb}{n^2N}.$$

This charge is released in an interval equal to that required by the beam

in traversing the charged element; namely, $t_e = b/n^2N$. Therefore, the instantaneous current is:

$$i_e = \frac{q}{t_e} = IAp.$$

Since each element behaves in this way, the average signal current is equal to this value. Assuming as before a 20,000-ohm coupling resistance, the signal-to-noise ratio, for an incident illumination of 10,000 foot-candles, and a mosaic whose area and photosensitivity are 20 square inches and 20 microamperes per lumen, respectively, is:

$$\frac{e_s}{e_n}\text{ (rms)} = \frac{IpAR}{3.5 \times 10^{-5}} \cong 5 \times 10^5.$$

In other words, an incident illumination of about 1 foot-candle would be needed to give the required signal-to-noise ratio. This estimate is, of course, based on an ideal tube, and does not take into account various inefficiencies occurring in an actual tube.

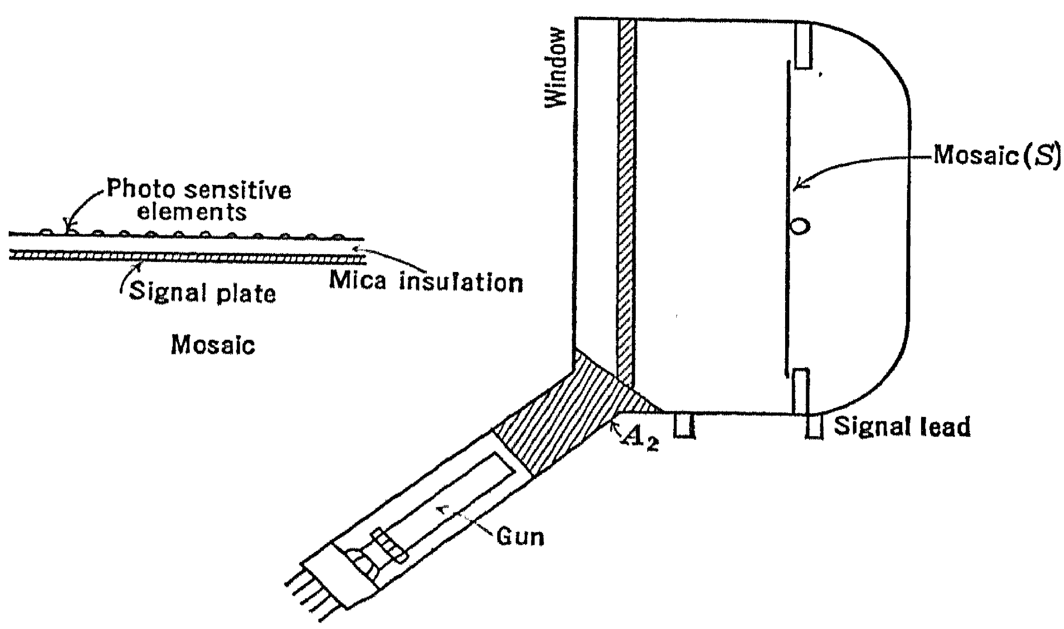


FIG. 8.20.—Normal Iconoscope with a Single-Sided Mosaic.

8.9. Iconoscope; Single-Sided Mosaic Type. The Iconoscope as used in practice operates on the same principle but has a somewhat different embodiment. The tube is diagrammed in Fig. 8.20. Instead of the two-sided mosaic built up on a mesh base, as in the preceding tube, the mosaic consists of a very thin mica sheet coated with conducting film on the side away from the gun, and with a myriad of tiny photosensitized silver elements on the other. Each of these elements is insulated from the others, and as a result of incident light stores a charge due to the emit-

ted photoelectrons. The scanning beam returns the elements to their equilibrium potential with respect to the collector A_2, just as in the tube using the two-sided mosaic. The actual cycle of operation which is described in detail in Chapter 10 is somewhat more complicated than that outlined above. This is because secondary electrons not only go to collector A_2, but also return to other parts of the mosaic.

The efficiency of this type of tube is fairly low, that is, only 5 or 10 per cent, as compared to an ideal tube. This is in a large measure due to the fact that the photoemission from the mosaic is unsaturated. Such a condition is to be expected since the potential of the mosaic element can never be greatly different from that required for secondary-emission equilibrium with respect to collector A_2. However, the photoemission, particularly for small light intensities, is more efficient than might at first be expected because of the effect of the secondary-emission electrons which return to the mosaic and tend to drive elements not directly under the beam slightly negative. A detailed analysis must also take into account the short-circuiting effect of these same returning electrons. The reason for the use of this type of mosaic instead of the two-sided form which permits saturated photoemission is purely technological. The difficulties encountered in trying to construct a two-sided mosaic suitable for the reproduction of a high-definition picture are enormous and present a serious obstacle to their practical production.

In spite of its inefficiency, the advantage gained by the use of the storage principle is so great that a tube of the type illustrated in Fig. 8.20 gives excellent performance, both for direct and film pickup. The experimentally determined sensitivity is sufficient so that a tube having a mosaic area of 100 cm^2 and working with an $f/2.7$ lens will transmit a scene illuminated by from 50 to 100 foot-candles with a signal-to-noise ratio of greater than 30, and will give a picture which is fully recognizable with less than 10 foot-candles illumination.

With all its high sensitivity, the Iconoscope is not without its faults. One of the most serious is a spurious signal which is the result of the unsaturated secondary emission required in its operation. This spurious signal takes the form of shading over the reproduced picture and must be compensated for by special circuit networks in the video amplifiers immediately following the pickup tube.

Various new types of pickup tube based on the Iconoscope principle are being investigated in laboratories all over the world. The aim of these new Iconoscopes is to improve the sensitivity and reproduction quality of the pickup device. A number of lines of attack are being adopted. For example, secondary-emission multiplication can be used to intensify the charge image received by the mosaic. This principle has

been successfully incorporated in the Image Iconoscope, a tube developed to a point where it has practical value.

Low-velocity scanning, barrier-grid mosaics, and similar expedients are the subject of active research as means for improving the operation of the mosaic itself, thereby eliminating the spurious signals and increasing the efficiency. Noise which determines the limit of sensitivity can be reduced by means of secondary-emission multipliers. These experimental tubes, together with many others, will be considered again in greater detail in Chapter 11.

The evidence at present indicates that electronic, storage pickup devices will, in the fairly near future, supplant all other systems. Mechani-

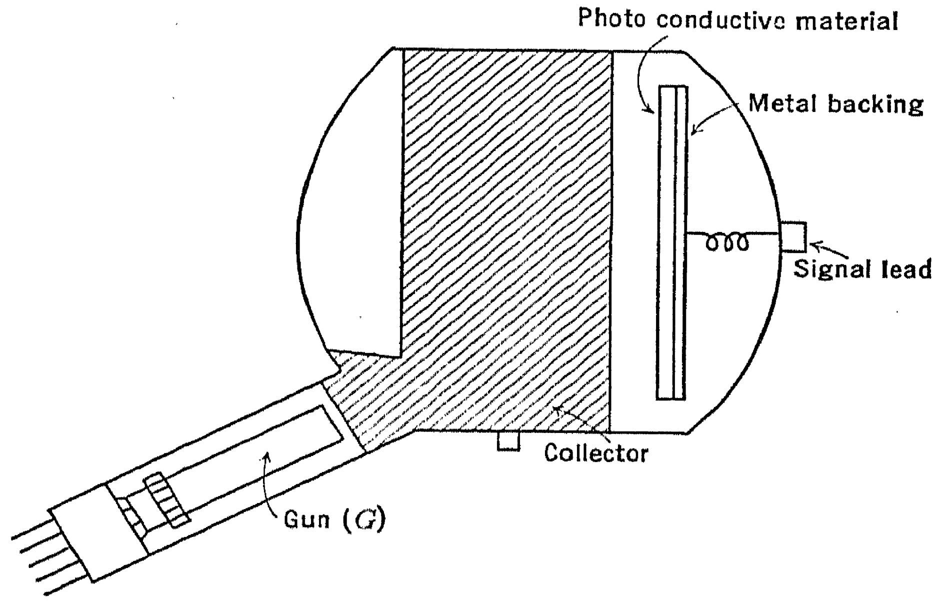


FIG. 8.21.—Photoconductive Pickup Tube.

cal systems capable of giving the required definition have been made and excellent results are obtained from them, but these devices represent almost the ultimate that can be attained in the relatively well-developed field of mechanics, while electronic systems, which even now excel mechanical pickup devices in most respects, rest on the relatively infant sciences of electronics and electron optics, and have almost unlimited possibilities of development and improvement.

8.10. Photoconductive Pickup Tube. Another class of electronic pickup devices, which is similar in appearance to the Iconoscope shown in Fig. 8.20, but quite different in operation, takes advantage of the photovoltaic and photoconductive properties of various materials. Fig. 8.21 illustrates a tube of this type, consisting of an electron gun producing a beam which sweeps over a thin surface layer of photoactive material. A potential is applied between the metallic backing and the electrode

collecting the secondary emission. The magnitude of this potential depends upon the particular tube, but usually it is somewhat less than that required to saturate the secondary emission so that the amount of current collected by A_2 varies with the electrical properties of the screen material at the point under the beam, these electrical properties being in turn dependent upon the light intensities. Copper oxide, selenium, etc., have been successfully used as screen materials in these tubes. Pickup tubes of this type are not in the strict sense of the word storage tubes, nor do they have all the characteristics of a non-storage system. Tests indicate that the fundamental sensitivity of these tubes is greater than that of an equivalent non-storage pickup arrangement. This is probably due to the fact that the accumulative effect so frequently present in photoconductive action is somewhat equivalent to storage.

8.11. Velocity Modulation. Before this survey of pickup devices is closed, another system based on fundamentally different television principles from those considered heretofore must be described. This system utilizes "velocity modulation." All the systems discussed in the preceding two chapters employ a scanning pattern which is independent of the contents of the picture transmitted. Picture reconstruction is effected by varying the brightness of the spot moving over this pattern. With velocity modulation the brightness of the reproducing spot remains constant, and variation in light intensity over the picture is obtained by varying the velocity of the spot as it moves over the viewing screen. Over bright areas of the picture the velocity of the spot is low, while low light intensities are produced by an extremely rapid motion of the spot. It is quite obvious that, since there must be a geometrical correspondence in the positions of the exploring spots of the transmitter and receiver, both spots must vary in velocity in the same way, that is, the deflecting means of the pickup and reproducing devices must be tied together by the communication channel. The fact that the scanning velocity must be altered rapidly almost completely excludes all mechanical systems from this type of transmission. However, any of the cathode-ray devices can be made to serve. The advantages of this system are twofold; first, theoretically at least, all problems of synchronizing are removed; and second, the beam at the reproducer is seen at maximum intensity at all times, thus giving a high screen brightness. However, the contrast range is rather limited because it is not possible to move the scanning beam fast enough greatly to reduce the light on the viewing screen. Furthermore, the resolution is not uniform over the picture but is greater in bright areas than in dark. There are also other practical drawbacks; for example, an amplitude distortion in the transmitting or reproducing units results

in a geometrical distortion of the image in addition to false brightness values.

An Iconoscope and associated circuits, arranged for pickup as a velocity modulator, are shown in Fig. 8.22. A video amplifier is connected to the signal plate of the Iconoscope in the usual way, and the output is supplied to the transmitting unit. In addition, the signal is also applied to the grid of the amplifier tube T_1 in such polarity that the signal corresponding to a dark area makes the grid of this tube more positive, thus increasing plate current. The horizontal deflecting plates of the Iconoscope are connected across the condenser C_1 so that the deflection of the beam is proportional to the charge in the capacitor. This condenser is charged by the plate current of T_1; therefore, the rate of deflection, i.e.,

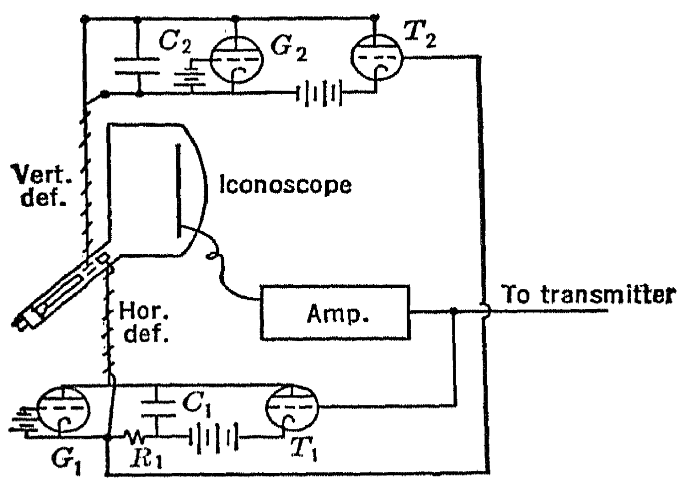


Fig. 8.22.—Velocity Modulation Pickup Using an Iconoscope.

velocity of the spot, is proportional to the signal, being great for the signal from a dark area and small for one corresponding to a bright area. The thyratron G_1 is also placed across the condenser, and, when its voltage exceeds a predetermined value, this tube becomes conductive, discharging C_1 and returning the beam to its initial horizontal position. The large discharge current flowing through the resistance R_1 drives the grid of T_2 momentarily positive, allowing a current pulse of definite magnitude to flow into C_2. The condenser C_2 is connected to the vertical deflecting plates, and its value is so chosen that each current pulse produces a vertical displacement of the beam equal to one-line spacing. When the charge in C_2 reaches a certain value it is in turn discharged by the thyratron G_2. It is interesting to note that, if the mosaic of the Iconoscope is fluorescent under the beam, a luminous image of the picture being transmitted will be formed.

The scanning spot at the receiver must merely follow the motion of the spot at the transmitter. By using deflecting circuits similar to those

at the pickup, and supplying the control tube with the video signal, this motion can be obtained. Fig. 8.23 shows the elements of the circuit for the receiver. The horizontal deflecting circuit differs from that of Fig.

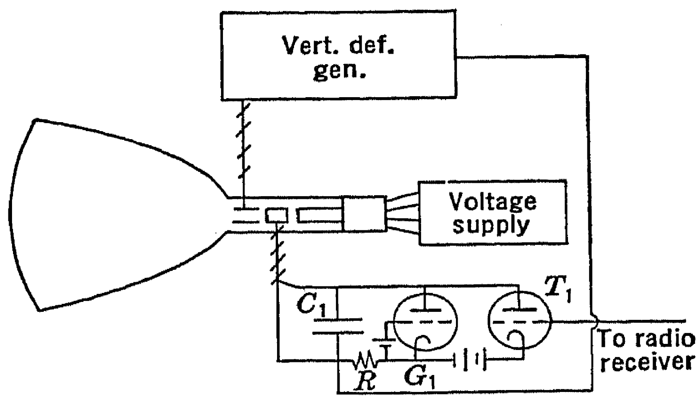


Fig. 8.23.—Reproducer of a Velocity Modulation Television System.

8.22 only in that the discharge of the condenser C_1 is triggered off by the signal from the transmitter. A special signal may also be required to generate the vertical return.

The video signal generated by the transmitter differs from that produced by intensity modulation. However, the bandwidth required for a

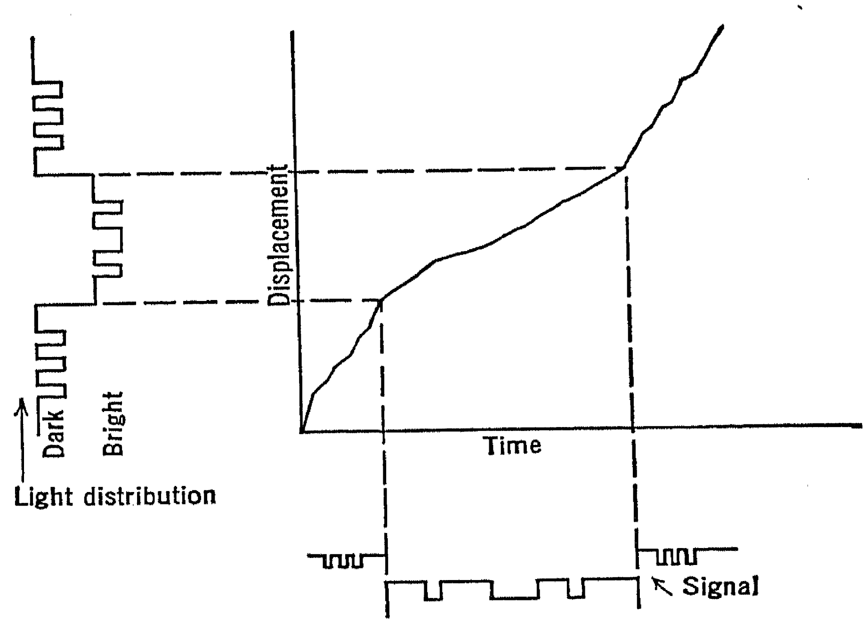


Fig. 8.24.—Video Signal Obtained with Velocity Modulation.

given average resolution is about the same for either system. The actual form of the signal and its spectrum can be determined by means of Fourier analysis, as was done in Chapter 6 for intensity modulation. Fig. 8.24 shows the signal for a few simple light-intensity distributions.

Velocity modulation is not a newcomer in the field. As early as 1911,

Rosing suggested it as a means of varying the light distribution on the screen of a cathode-ray viewing tube, although he failed to realize that the scanning at the transmitter has to be similar to that at the receiver if geometrical distortion is to be avoided. Before the general use of cathode-ray pickup tubes, this type of transmission proved successful as a means of pickup from moving-picture film.* The arrangement by which this was accomplished is shown in Fig. 8.25. The spot from a Braun tube was imaged on the film to be transmitted. The light passing through the film was picked up by a phototube. The output of this tube was amplified and supplied to the control tube (equivalent to T_1, Fig. 8.22) of the horizontal deflection. Therefore, the velocity of the light spot scanning the film is a function of density distribution of the film.

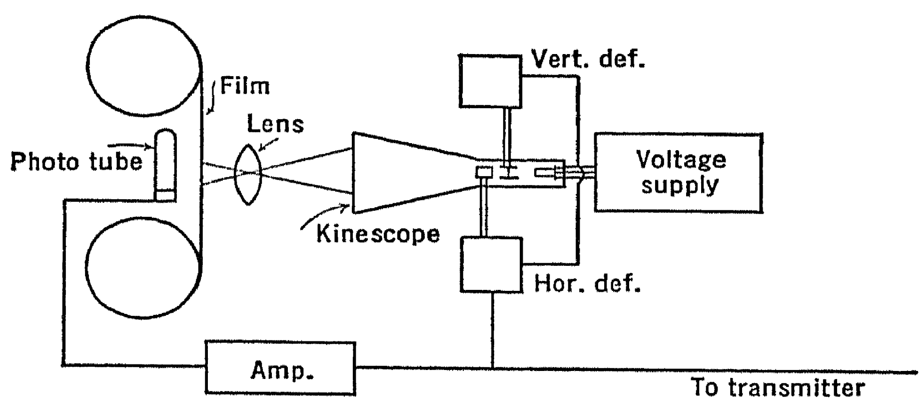


FIG. 8.25.—Flying Spot Film Scanner for Velocity Modulation.

The picture being transmitted is, of course, reproduced on the screen of the cathode-ray tube which furnishes the scanning light spot, as well as on the screen of the viewing tube at the receiver.

As has been pointed out, a high contrast ratio cannot be obtained by this system of transmission. To overcome this deficiency, experimental pickup systems combining velocity and intensity modulation, using the latter to reduce the scanning beam current in the viewing tube in order to produce black areas, have been set up. This does not greatly reduce the gain in brightness characteristic of velocity modulation, while permitting high values of contrast.

In spite of its theoretical interest, velocity modulation has not been considered of much practical importance, the gain in brightness at the viewing tube being more than offset by the added complications in equipment, the non-uniformity of resolution, and the difficulty in avoiding geometric distortions due to non-linear channel characteristics or interference.

This chapter is intended to outline a few of the more important pickup

* See Bedford and Puckle, reference 12.

systems. Many pickup systems which, though different in many important respects, are similar in basic principles were unavoidably omitted. Some of these systems have given very good service in experimental television tests.

Because of the very great theoretical and demonstrated practical advantage of the storage-type pickup device, particularly in the field of direct pickup, its construction, theory, and performance must be examined in more detail. This further consideration will be reserved for Chapters 10 and 11.

REFERENCES

1. A. Dinsdale, "First Principles of Television," John Wiley, New York, 1932.
2. J. H. Reyner, "Television, Theory and Practice," Chapman and Hall, London, 1934.
3. F. Schroeter (editor), "Handbuch der Bildtelegraphie und des Fernsehens," Julius Springer, Berlin, 1932.
4. J. C. Wilson, "Television Engineering," Pitman, London, 1937.
5. F. Schroeter, "Fernsehen," Julius Springer, Berlin, 1937.
6. George Newnes, Ltd., "Television Today," Vols. I and II, London, 1936.
7. A. A. Campbell Swinton, "The Possibilities of Television," *Wireless World and Radio Review*, pp. 51–56, April 9, 1924.
8. P. T. Farnsworth, "Television by Electron Image Scanning," *J. Franklin Inst.*, Vol. 218, pp. 411–444, October, 1934.
9. V. K. Zworykin, "The Iconoscope," *Proc. I. R. E.*, Vol. 22, pp. 16–32, January, 1934.
10. V. K. Zworykin, G. A. Morton, and L. E. Flory, "Theory and Performance of the Iconoscope," *Proc. I. R. E.*, Vol. 25, pp. 1071–1092, August, 1937.
11. H. Iams and A. Rose, "Television Pickup Tubes," *Proc. I. R. E.*, Vol. 25, pp. 1048–1070, August, 1937.
12. L. H. Bedford and O. S. Puckle, "A Velocity Modulation Television System," *J. Inst. Elec. Eng.*, Vol. 75, pp. 63–82, July, 1934.

CHAPTER 9

PICTURE REPRODUCING SYSTEMS

9.1. Mechanical Scanning. The division of pickup devices into two classes has its counterpart in a similar subdivision of viewing devices, classifying them into mechanical and electronic systems. Because of its historical priority the class of mechanical systems will be considered first.

The principal components of a mechanical reproducer are the scanning mechanism, the light source, and the means for controlling the light intensity. In the earliest systems the last two elements were combined, that is, the light source was a gaseous glow discharge and its brightness was varied by exciting it directly with the video signal. A Nipkow disk with a spiral row of apertures equal in number to the number of lines making up the picture effected the scanning. The glow lamp was designed in such a way that the discharge was spread uniformly over an area equal to the picture area. This made possible the very simple arrangement shown in Fig. 9.1. Later the use of a crater lamp and suitable optical system greatly increased the brightness of this type of reproducer.

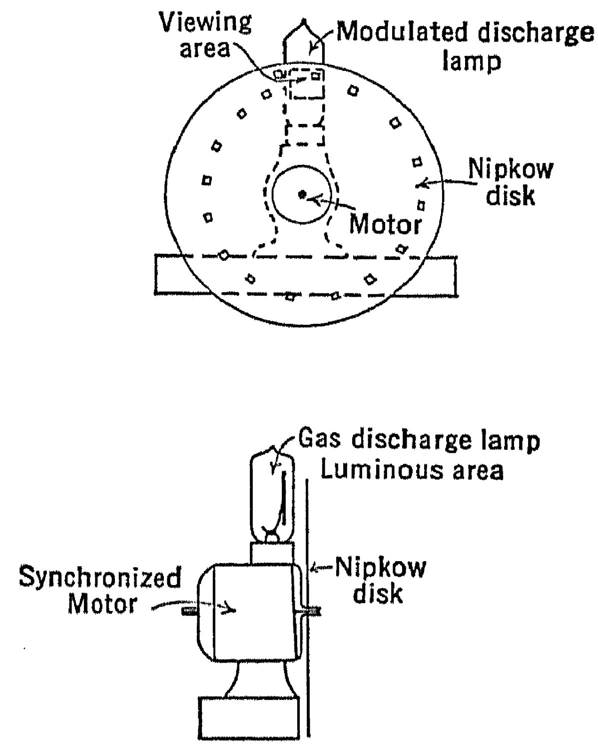


FIG. 9.1.—Simple Nipkow Disk Viewer.

The frequency response of the glow-discharge type of light source is very much too limited to make this arrangement feasible for high-definition purposes. Furthermore, when a large number of lines are used the amount of light which can be obtained through a simple aperture disk, even with a very intense arc source, is very small, as has already been pointed out.

In order to meet the difficulty resulting from the limited light passed by the scanning disk, a number of optically more efficient scanners are possible. The lens disk and lens drum improve the optical efficiency, but they are both expensive to make and critical in adjustment. Mirror and

prism drums or screws have been found to meet the necessary optical requirements without the disadvantages of the former. Several alternative forms of the simple mirror drum are shown in Fig. 9.2. Each of the drums shown has segments equal in number to the number of lines in the picture and is driven at frame frequency. Unless the drum is very large indeed, the mirror segments for high-definition work must be very narrow. Consequently, some increase in optical efficiency can be ob-

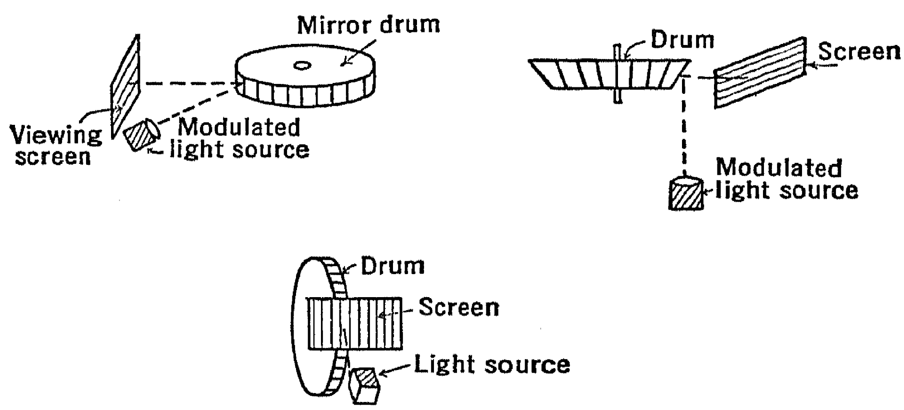


Fig. 9.2.—Modifications of the Mirror Drum.

tained by introducing combinations of cylindrical lenses instead of the simple spherical lens shown. This lens or combination of lenses images an aperture at the light source onto the viewing screen.

An interesting variation on the mirror drum is shown in Fig. 9.3. The segmented drum is stationary and is made up of internally placed mirrors. A two-sided plane mirror in the center revolves at frame frequency

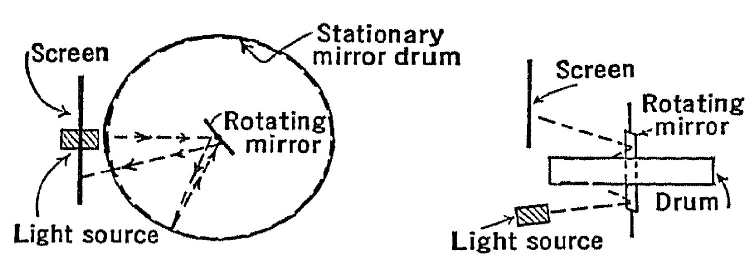


Fig. 9.3.—Scanning System Employing a Stationary Mirror Drum (Mihaly).

and produces the scanning action. The size of the mirror drum becomes very large for anything approaching high-definition work. To effect a reduction in size a polyhedral mirror can be substituted for the plane revolving mirror.

The mirror screw is another mechanical scanner for picture reproduction. An array of striplike mirrors is arranged along a shaft in such a way that their narrow dimension includes the axis of the shaft and their faces are each inclined at a small fixed angle with respect to the preceding

mirror. In appearance the assemblage resembles a screw, thus giving the arrangement its name. Fig. 9.4 shows diagrammatically such a scanner. The number of mirrors is equal to the number of lines used, and the whole structure is rotated at frame frequency. The picture is viewed directly on the screw, so that the observer sees the image in space, apparently located behind the scanner, since it is made up of the direct reflection of the modulated light source as seen in the spinning mirror screw.

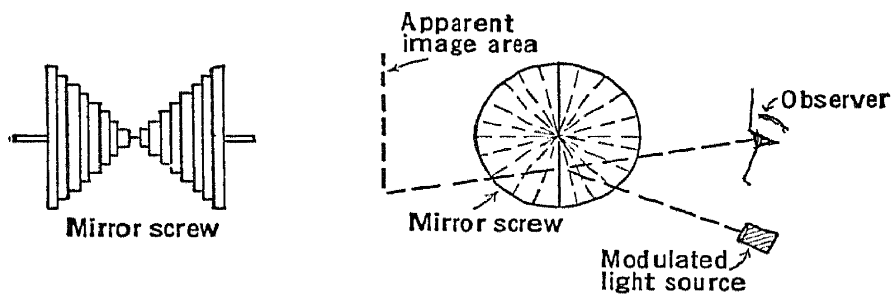


FIG. 9.4.—Mirror Screw.

Another modification of the mirror drum principle involves two drums, one rotating at high speed to give line scanning, the other at a much lower speed to produce the necessary displacement of the line. When cylindrical lenses are employed this system can be made optically quite efficient. The two drums make it possible so to reduce the size of

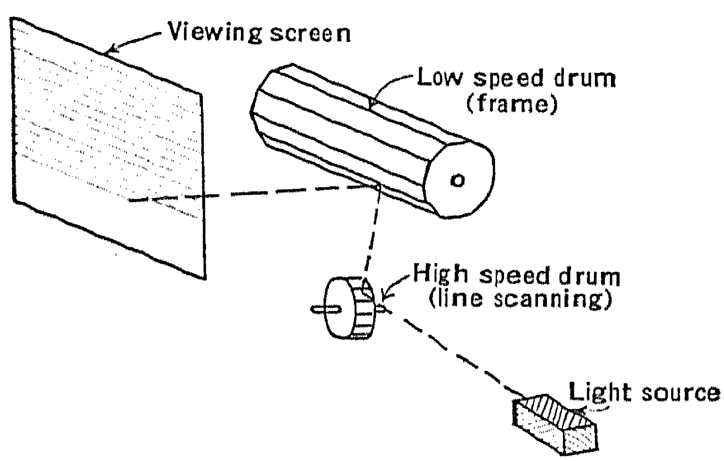


FIG. 9.5.—Television Reproducer Using High- and Low-Speed Mirror Drums.

the equipment that it is no longer totally impractical for high-definition work. A schematic diagram of the arrangement is shown in Fig. 9.5.

9.2. The Kerr Cell. One of the major problems in connection with mechanical viewing systems is the modulation of a light beam of sufficient intensity to give a usable picture. The direct modulation of a gas discharge or arc is immediately excluded, because no known source of this type can respond to variations of the frequency necessary for high-

definition pictures. Mechanical light valves (excluding supersonic wave motion in elastic media), such as oscillating mirrors or shutters, suffer from the same limitation. Even the most nearly inertia-free systems of strings or reflectors will not respond to frequencies much above 50 kilocycles, whereas high definition requires nearly 100 times this frequency.

There are several electro-optical effects which suggest themselves as a means of performing this function. Upon investigation, only the Kerr effect has been found to offer anything resembling a practical solution. Therefore, the discussion of electro-optical light valves will be limited to those based on this phenomenon.

Certain organic liquids, when subjected to an electric field, lose their isotropic optical properties and become birefringent. This is known as the Kerr effect. When a beam of polarized light passes through a birefringent material, the velocity with which it traverses the substance

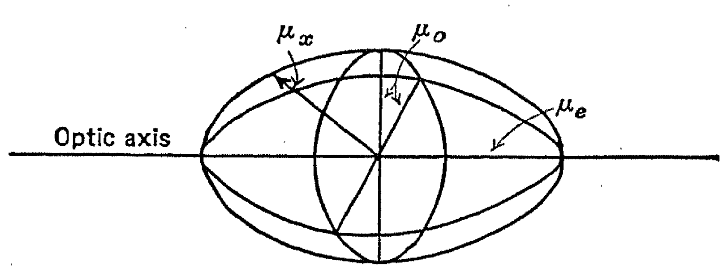


FIG. 9.6.—Ellipsoid of Refractive Index for a Uniaxial Birefringent Material.

depends upon the direction of the electric vector of the incident light. In other words, the index of refraction of the material depends upon the direction of propagation and plane of polarization of the light. The index of refraction may be represented as an ellipsoid, as shown in Fig. 9.6. If this ellipsoid is one of revolution, there is one direction through the material for which the index is independent of the plane of polarization. This is known as the optic axis, and the material may be classified as uniaxial. Substances which exhibit the Kerr effect are uniaxially birefringent under the influence of the field, with the optic axis in the direction of the field.

The mechanism of the effect may be pictured as follows: The molecules of the material (a liquid usually) are polar, that is, each molecule contains an equal positive and negative charge which can be displaced relatively in some preferential direction. Along the direction in which this displacement can take place, the effective "dielectric constant" is greater than in any other direction. In the absence of a strong constant electric field, the molecules are randomly oriented and the medium is isotropic. An electric field causes a preferred orientation to be super-

imposed on the statistical randomness of the molecules. This makes the material anisotropic.

If a polarized beam of light passes through an isotropic medium it remains unchanged, both in degree and plane of polarization. For a birefringent material this is not true. A beam of light, polarized in a plane at 45° with respect to the optic axis and directed normal to the axis, is divided into two components, each of which traverses the medium at slightly different velocity. Upon emerging the two components will, in general, differ in phase, and the light will no longer be plane polarized. If, however, the delay between the two components is half a wavelength, the emerging light will be polarized but in a plane normal to the original plane. The extent of the delay depends upon the medium, the applied voltage, and the length of path. This difference expressed as a phase shift is:

$$\phi = 2\pi B l E^2,$$

where E is the field strength in the liquid, l the length of the light path, and B a constant, depending upon the medium, known as the Kerr constant.

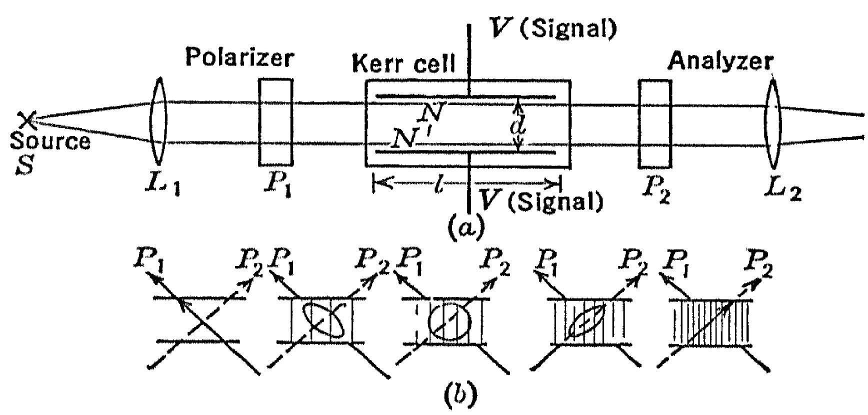


FIG. 9.7.—Principle of the Kerr Cell Light Valve.

One way of using this effect in a light valve is shown in Fig. 9.7*a*. Light from the source S is polarized by the element P_1, which may be a Nicol prism, a Polaroid sheet, or similar means, in a plane making an angle of 45° to the electrodes NN' in the cell. The element P_2 is the analyzer and is similar in construction to the polarizer P_1. It passes light polarized in a plane at right angles to that of the light emerging from P_1. The heart of the valve is the cell, which is filled with a liquid exhibiting the Kerr effect and contains two plane parallel electrodes NN' to which the modulating voltage V is applied.

When no potential is applied to the electrodes, a negligible amount of light passes through the system since in effect it is one consisting of

crossed polarizers. However, when a potential is applied between the plates, making the liquid anisotropic, light can pass through the analyzer. The amount of light increases with potential until, when the light is completely polarized in the plane of transmission of the analyzer P_2, half the light entering will be transmitted (half being rejected by the polarizer P_1). A further increase in potential again decreases the light. The light emerging from the cell is represented diagrammatically in Fig. 9.7*b*. Quantitatively, the amount of light emerging I is given by:

$$I = \frac{I_0}{2} \sin^2 \frac{\phi}{2},$$

where I_0 is the light entering the system. This, in terms of the cell constants, is:

$$\begin{aligned} I &= \frac{I_0}{2} \sin^2 \pi \frac{BlV^2}{d^2} \\ &= \frac{I_0}{4}\left(1 - \cos 2\pi \frac{BlV^2}{d^2}\right), \end{aligned}$$

where V is the potential between plates and d the spacing between them. This expression has its maximum value when:

$$V_{\text{max.}} = d\sqrt{\frac{1}{2Bl}}.$$

An idealized curve showing the variation in light with applied voltage

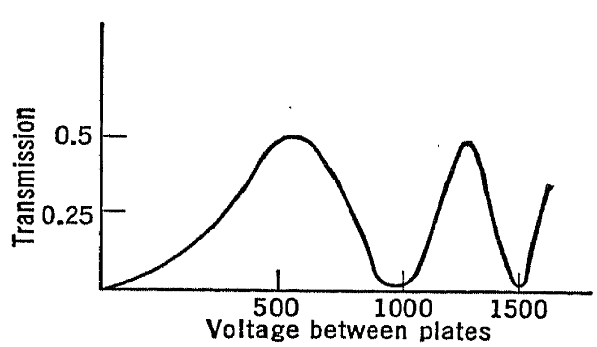


FIG. 9.8.—Light Transmission Characteristic of a Typical Kerr Cell.

is shown in Fig. 9.8. Unfortunately, the Kerr effect is independent of neither temperature nor wavelength of light. Therefore, not only must the cell be kept at constant temperature, but also some correction must be made for the dispersion of the cell. A mica halfwave plate can be made to serve the latter purpose.

The Kerr constant B for a number of liquids is given in Table 9.1. From the magnitude of the Kerr constant it is evident that, even with

TABLE 9.1

Substance		$B \times 10^6$(esu)	Resistivity	Dielectric Constant
Carbon disulphide	CS	0.36	13×10^{15}	2.6
Chloroform	$CHCl_3$	0.32	5×10^7	5.0
Acetone	C_3H_6O	1.6	8×10^6	21.0
Nitrobenzene (purified)	$C_6H_5NO_2$	41.0	10^{10}	36.0

a nitrobenzene cell, the spacing between the plates must be very small or the voltage needed across the plates will be enormous. For example, a cell 5 cm long will have a maximum voltage required for the change from complete extinction to maximum transmission:

$$V = d\sqrt{\frac{10^6}{2 \cdot 40 \cdot 5}} \times 300.$$

It can be seen that, with d = 1.0 mm, about 1500 volts are required. If in operation the cell is biased to 1100 volts, the peak control voltage is some 400 or 500 volts. The narrow spacing and great length of the cell mean that the optical aperture of the system will be very seriously limited. To overcome this, Karolus suggested a multiple cell. A cell of this type is illustrated in Fig. 9.9. The capacity, together with the high control voltage, makes this cell rather impracticable because of the very great power required to drive it. Power at video frequency is both difficult to obtain and expensive. In fact, this fault excludes the multiplate cell from practical high-definition work.

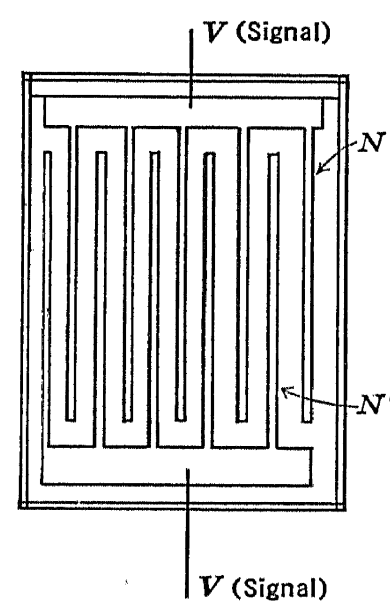


FIG. 9.9.—Karolus Multiple Plate Kerr Cell.

The Wright cell is a somewhat more practical solution. Fig. 9.10 illustrates this type of cell. If an image of the scanning aperture is focused on the center of the cell, the system may be made fairly efficient optically.

As has been pointed out, half of the light is lost at the first polarizer. Several systems have been suggested to avoid this loss of light. Only one of these will be discussed. The arrangement is shown in Fig. 9.11. Instead of the initial polarizer a calcite crystal divides the incoming light

into two parts polarized at right angles and slightly displaced, owing to the two indices of refraction. A second calcite replaces the analyzer. This still further separates the two beams, so that they fall onto the mask on either side of the slit. If the Kerr cell between the plates causes a shift of 90° in the plane of polarization of each ray, the two rays will

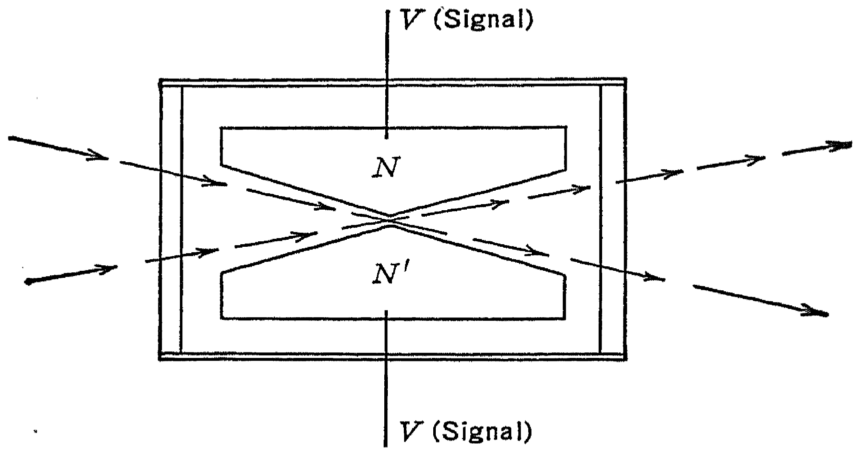


FIG. 9.10.—Wright Wedge Kerr Cell.

be bent in the second crystal in such a way that they meet upon emerging and pass through the slit. At lower control voltages, when the cell only elliptically polarizes the light, only a fraction will pass through the slit.

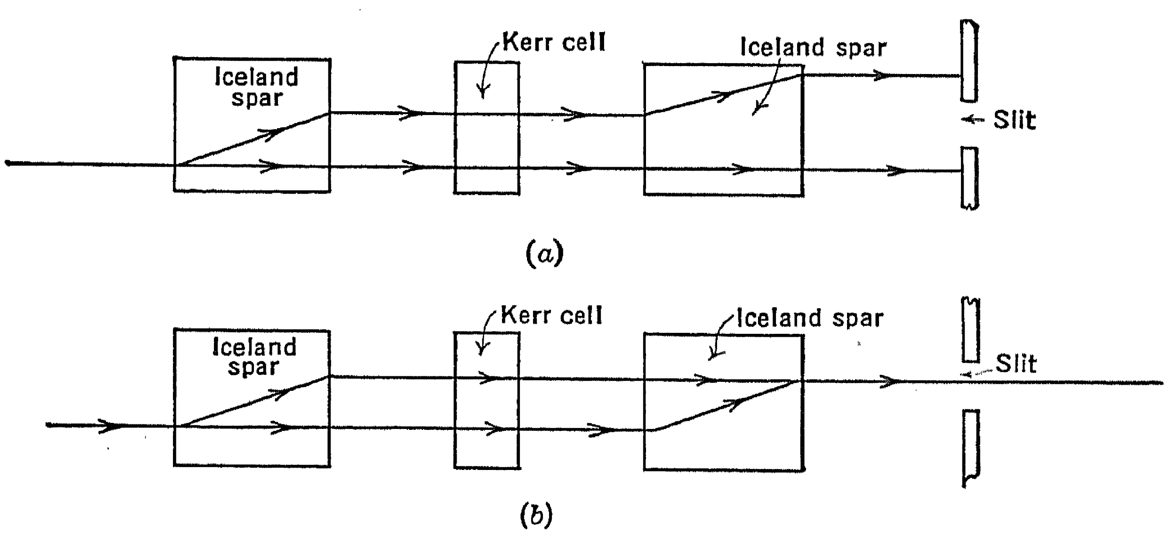


FIG. 9.11.—Modified Kerr Cell Light Valve Utilizing Nearly All Incident Light.

If there were no alternatives, the Kerr cell would offer a possible means of obtaining a high definition picture. However, the power required from the video system, the limitations it imposes upon the optical arrangement, together with other less important objections such as nonlinearity of control, render the system less suited to the present re-

quirements than other arrangements, in particular, electronic viewing devices.

9.3. Supersonic Light Valve. Another, radically different, light valve is also used in television reproduction. The action of this valve is based upon the diffraction of light by supersonic waves in a transparent medium. The method of exciting supersonic waves in liquids by a piezoelectric crystal was developed by Langevin about 1917, and, furthermore, Brillouin at an early date had considered mathematically the scattering of light by thermal motion treated as superimposed elastic waves. It was not until 1932, however, that the diffraction of light by supersonic waves was observed experimentally. The phenomenon was shown independently by Debye and Sears in this country and by Lucas and Biquard in

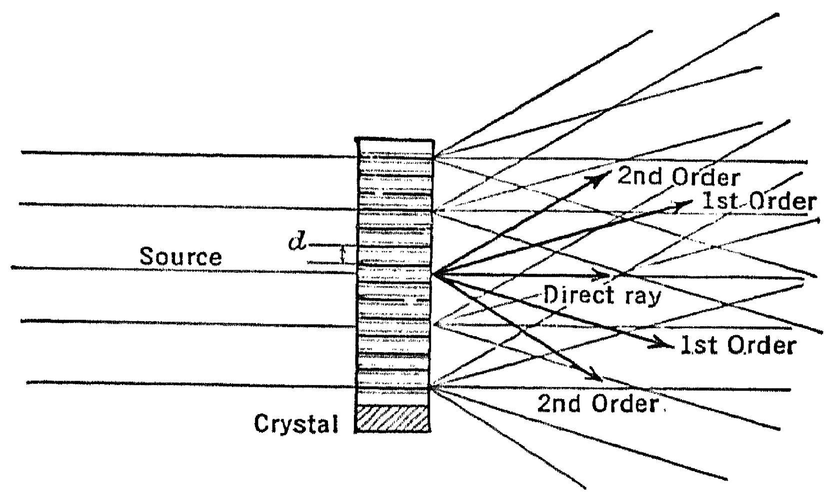


FIG. 9.12.—Diffraction of Light by Supersonic Waves.

France. Their observations indicated that, if plane waves were excited in a trough containing a liquid such as carbon tetrachloride, benzine, kerosene, or glycerine, a light beam parallel to the wave fronts would be efficiently diffracted. The experimental arrangement is shown in Fig. 9.12. The waves in the liquid are made up of alternately compressed and rarefied regions, and consequently the index of refraction of the liquid varies sinusoidally in correspondence with these compression waves. If the velocity of propagation of the elastic wave is v, and the excitation frequency f, the separation between successive high-index regions (i.e., wavelength) is $v/f = d$. The light which passes through the regions of high index is delayed slightly more than that through the lower-index portions, so that the emerging light has a periodically varying phase normal to its direction of propagation, or, considering planes of constant phase, these planes are corrugated. On the basis of this picture of the action of the cell it is possible, with the aid of Huygens' principle, to calculate the intensity and position of the diffracted beams and the in-

tensity of the direct or central beam. The positions of the diffracted beams are found to be given by the relation:

$$n\lambda = d \sin \theta,$$

where n is the order of the diffracted ray, λ the wavelength, and θ the angle measured from the direction of the central beam. The intensity of the light going into the diffracted rays depends upon the amplitude of the supersonic waves, and, with sufficient amplitude, practically all the light can be thrown from the central ray into the scattered beams. It is interesting to note that the position of the orders of diffracted light is unchanged by any change in amplitude of the elastic waves.

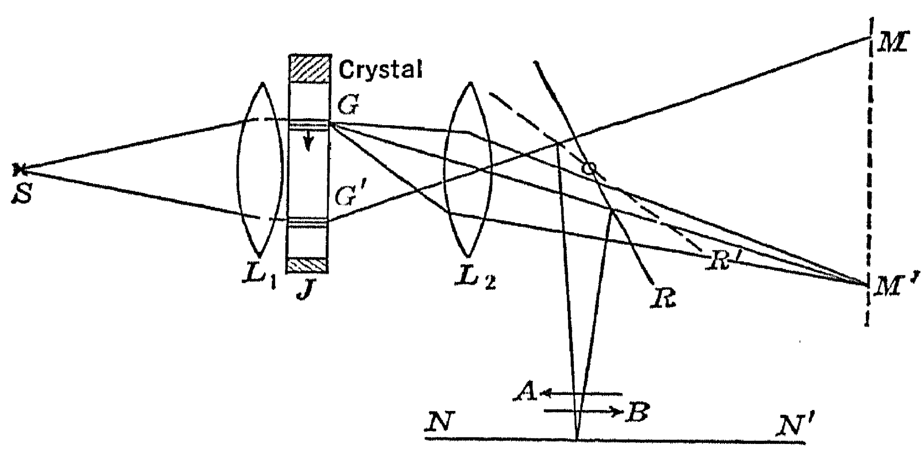


Fig. 9.13.—Optics of the Supersonic Light Valve; Method of Arresting the Motion of the Image of a Picture Element.

A very ingenious application of this effect to the problem of television reproduction was made by Jeffree in 1936. A supersonic light valve based on this principle forms the nucleus of the present Scophony system.* The cell consists of a fairly long trough filled with a suitable liquid and having a quartz piezoelectric crystal at one end. At the other end there is a layer of absorbing material such as glass wool so that the compression waves are not reflected back. When the crystal is made to oscillate, traveling waves move down the trough at a velocity v which depends upon the density and elastic constants of the liquid. In practice a liquid giving a velocity of about 1000 meters per second is used, while the normal frequency of the crystal is in the neighborhood of 10 megacycles, resulting in a wavelength of $10^5/10^7 = 0.01$ centimeter. A cell of this type is shown in Fig. 9.13.

If the crystal is driven for a few cycles, a wave group will move through the trough at the propagation velocity v. In the figure, G and G' indicate two positions of such a group as it moves up the trough. The trough is

* See Lee, reference 6.

illuminated from the source S, which in conjunction with the lens L_1 produces a parallel beam of light through the cell. Lens L_2 images the cell into the plane MM', but the mirror R intercepts the light from this lens and forms the actual image on the screen NN'. By an optical arrangement which will be explained below, only the light diffracted by a disturbed region is permitted to pass through the system. Therefore, if the mirror is held stationary the image of the wave group G will appear as a bright area moving across the screen NN' in the direction indicated by the arrow A. However, if the mirror rotates in such a direction that the image of a stationary point on the cell moves with equal velocity in the opposite direction across the screen, as indicated by the arrow B, the bright image of the wave group will appear stationary.

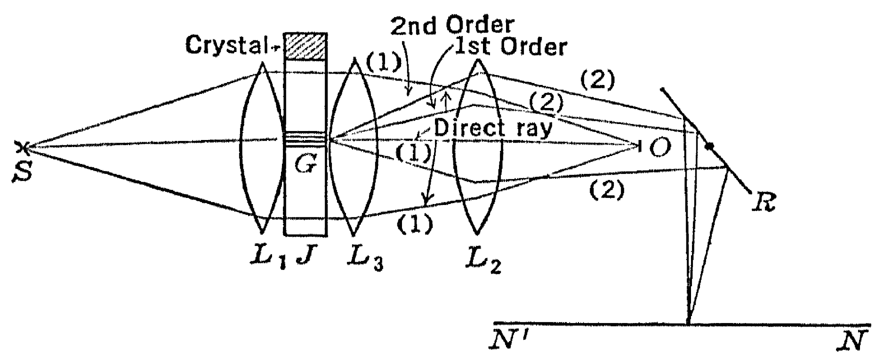


FIG. 9.14.—Optics of the Supersonic Light Valve; Method of Controlling Light Transmission.

In order to remove the direct beam and use only the diffracted light produced by periodic variations in the index of refraction of the liquid, the optical arrangement may be as shown in Fig. 9.14. The combination of lenses L_1 L_3 L_2 images the source S onto the shield O. Undisturbed rays 1 indicate this image formation. If the fluid is undisturbed, all the light entering the system is stopped by this shield, so that the screen is entirely in darkness. A disturbed region in the cell, such as G, causes some of the light to be thrown into diffracted rays. These rays (represented by rays 2 in the diagram) are not stopped by the shield, but instead are brought together by lenses L_3 and L_2 on the screen NN' described above.

Used in a television system an oscillator drives the crystal at resonance, usually, as has already been mentioned, at a frequency around 10 megacycles. The amplitude of this oscillation is modulated with the video signal in such a way that the signal corresponding to a black area leads to zero amplitude, while a bright area gives large amplitudes. If the instantaneous distribution of amplitudes in the cell is considered, it will be seen that they record the video signal for a period of time equal

to l/v, where l is the length of the cell. By making the length of the cell within the optical system equal to the product of the velocity of propagation and the line period, the distribution corresponding to a complete line will be included. Therefore, the optical system images a light distribution corresponding to an entire picture line on the screen. As the distribution moves through the cell, the light image is held stationary by the motion of the mirror. In an actual television reproducer the moving mirror takes the form of a small mirror drum, so that as one face carries the image off the screen an adjoining face brings another image onto the screen. A

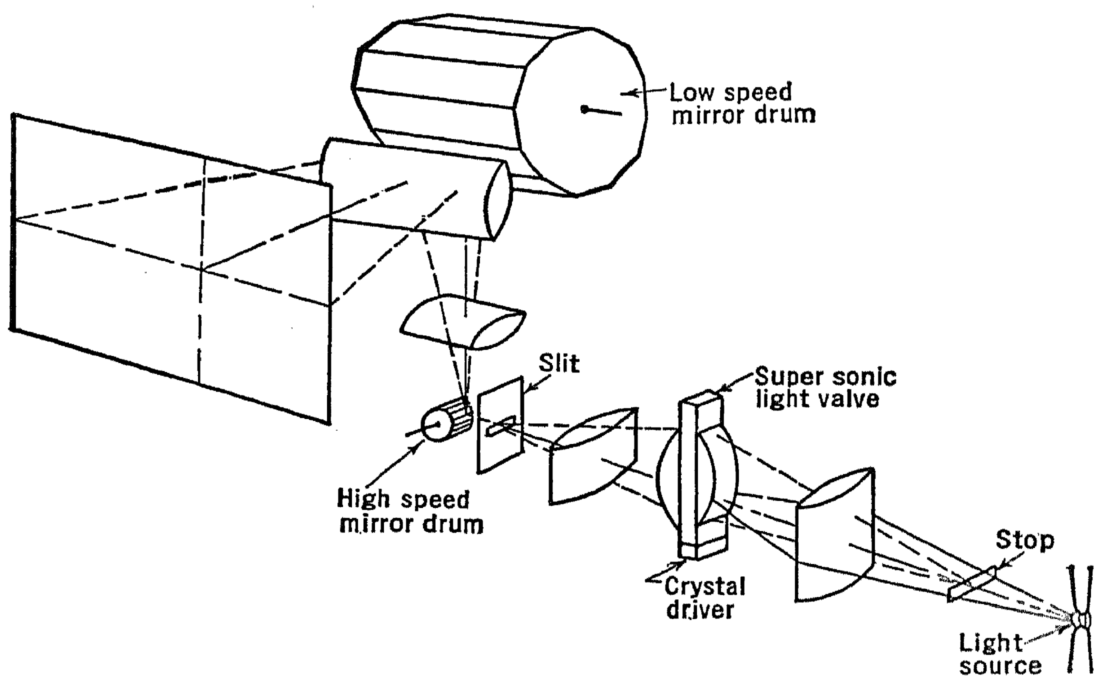


FIG. 9.15.—Schematic Diagram of a Scophony Television Receiver Employing a Supersonic Light Valve.

second mirror drum provides scanning motion at right angles to the direction of the lines.

The complete optical system of a reproducer using this system is shown in Fig. 9.15. In this arrangement the shield has been moved up to the light source and a slit is located at the position previously occupied by the shield. However, it will be seen that the action is the same as described above. The small line scanning mirror drum must run at a very high speed. For example, if this drum has 20 faces it must be driven at more than 39,000 rpm for a 441-line picture. Therefore, it must be very accurately balanced and have the minimum practical moment of inertia. It will be noticed that cylindrical lenses rather than the usual spherical lenses are used throughout the optical system. Such elements greatly

reduce the size of the equipment necessary to obtain a given optical aperture.

The supersonic light valve gives in one sense a storage of the information of the picture elements over a period of a line. However, it does not lead to the full light gain that would be expected from an unrestricted storage system. The limitation is due to the fact that light must pass through the cell in an essentially parallel beam. As a consequence of this limitation, it is evident that a light valve which has the same area as the supersonic cell, but which controls the transmission simultaneously over its entire area, can be made to produce the same illumination on the viewing screen. Such a system is illustrated diagrammatically in Fig. 9.16. Only one mirror drum is shown, although two are actually required. For the sake of economy of size in the moving parts, cylindrical lenses would be used. Similarly, a much smaller cell which would control light entering from all directions could be made to produce the same results.

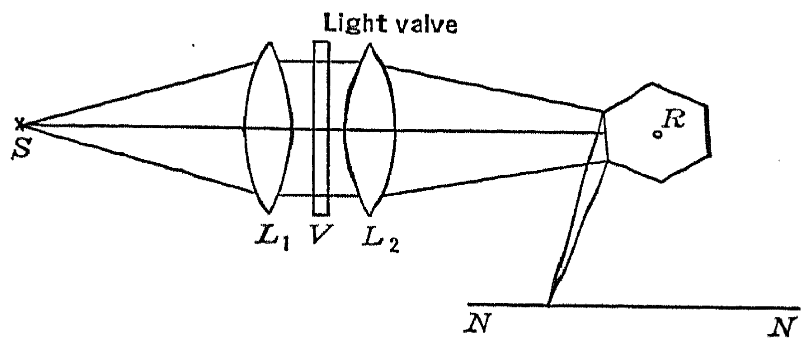


FIG. 9.16.—Large Area Light Valve.

However, in the absence of these ideal valves, the supersonic cell, in spite of its complications, has shown itself to be by far the most efficient arrangement to be used in connection with the mechanical scanner for the reproduction of television pictures. The amount of light that can be obtained is sufficient, when a high-intensity moving-picture type of arc serves as a light source, so that the image formed on a 5 by 6 foot screen can be viewed without discomfort.

9.4. Electronic Viewing Devices. In the field of image reproduction the purely electronic viewing device has had outstanding success. By its use all mechanical moving parts are eliminated and the problem of producing and controlling light of the required intensity is greatly simplified. Consequently the cathode-ray viewing tube is widely used today in practical television receivers. The history of the developments leading up to the modern cathode-ray tube is one with that of electronics itself, and as such would require many pages to present even in outline form. Therefore, it must be omitted from this necessarily brief account.

The first proposal to use a cathode-ray tube as a means of reproducing

television images appears to have been made by Boris Rosing, a year or two before Campbell Swinton's suggestion of applying cathode rays to the problem of pickup. Before the end of the first decade of the twentieth century, Rosing was able to demonstrate the synthesis of simple images by means of a Braun tube. The Braun tube is a modification of the Crookes discharge tube, arranged as shown in Fig. 9.17, to produce a narrow pencil of cathode rays which can be deflected by means of electrostatic plates or electromagnetic coils to any point on the fluorescent screen forming the front wall of the tube. The electrons forming the beam in this tube are produced by the ionization of gas in the vicinity of the cathode. An aperture defines the beam while another aperture, in conjunction with the deflection plates, serves to control the current reaching the screen. The scanning pattern is generated by supplying sawtooth current waves of suitable periodicity to the two mutually perpendicular deflecting coils. This relatively primitive cathode-ray tube would not,

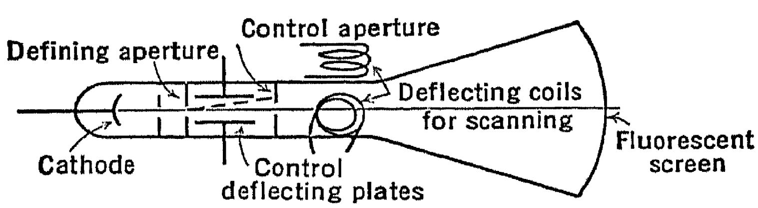


Fig. 9.17.—Rosing's Electronic Viewing Tube.

of course, fulfill present-day requirements. The fine spot required for 441 lines could not be obtained, control of the beam current is inadequate, and high deflection frequencies are impossible with this type of gas discharge. However, it served the need of the moment.

As the technique of picture transmission advanced and the standards of requirements were raised, greater demands were put upon the viewing tube. However, the improvement of the cathode-ray tube more than kept pace with these developments. Today it is capable of reproducing in black and white a 441-line picture suitable either for direct viewing or for projection on a large screen. Furthermore, it can be stated with assurance that the improvement in performance will continue to keep step with the inevitable advance of television standards.

The essential elements of the cathode-ray tube as used for television reproduction are: a means for producing a fine pencil of electrons, usually known as the electron gun; an arrangement for deflecting the electron beam in two mutually perpendicular directions; and a screen coated with luminescent material. These elements are assembled in a glass bulb which is evacuated to as low a pressure as is practicable. A typical ex-

ample of a tube of this class is shown in Fig. 9.18. When the tube is operated the synthesized picture appears on the large screen forming the front of the bulb, either to be viewed directly or else projected onto a suitable large viewing surface. This picture has the line structure characteristic of the scanning process used in television. The lines are produced by the motion of the bright spot which results from the cathode-ray beam striking the fluorescent material of the screen. The motion is obtained by bending the electron stream close to the end of the gun. Either an electrostatic or magnetic field is suitable for producing the deflection. Two fields at right angles to the axis of the gun are employed in either case. These fields may be superimposed or may be sequential. The field strength of one is made to vary periodically at line frequency, while the other has a periodicity corresponding to frame (or field) frequency. The variation is such as to deflect the beam at a uniform rate

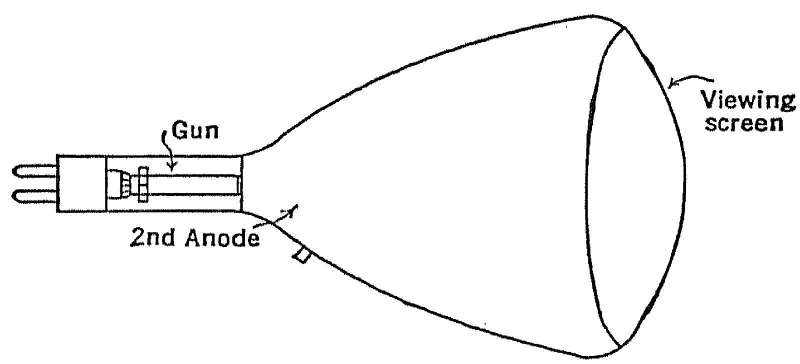


FIG. 9.18.—Schematic Diagram of Direct-Viewing Kinescope.

across the screen and return it rapidly to its original position. The superposition of the two displacements results in a parallel line scanning pattern of the type described in Chapter 6.

The brightness of the reproducing spot is varied by controlling the current in the scanning beam. This control must be effected without altering either position or size of the spot, and for practical reasons should not require large amounts of video power.

The importance of the electron gun is evident. It is responsible for supplying to the screen all the power which is needed to produce the light. In order to obtain this power not only must the electrons be accelerated to a high velocity in the gun, but also the beam current must be quite large. The beam voltage for the type of tube generally used in direct-viewing receivers is 6000 or 8000 volts; tubes used for projection work require 10,000 to 20,000 volts, or even higher voltages. The current needed varies from 100 to 500 microamperes in the former to several milliamperes in the latter. This power must be delivered into a very small spot if the resolution of the image reproduced on the screen is not

to be limited by the cathode-ray tube. For a 441-line picture the current must be concentrated into a spot which is only slightly greater than a 1/441 part of the vertical height of the picture in diameter—in other words, for a 12-inch direct-viewing tube, about 0.02 or 0.03 inch, while in a projection tube it may be as small as one-fifth of this value.

The gun consists of an electron source and an electron-optical arrange-

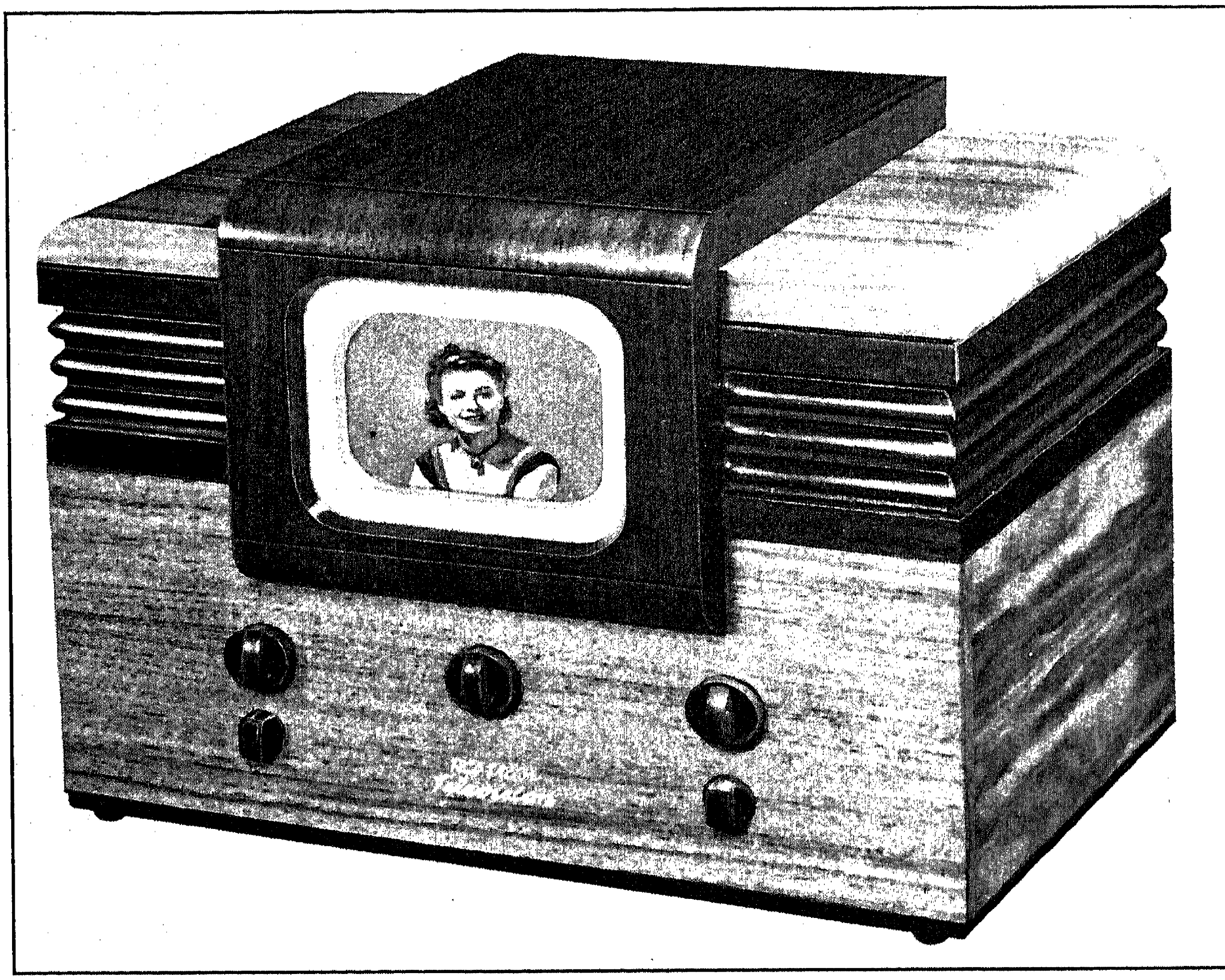


FIG. 9.19.—Table Model Television Receiver RCA-TT5.

ment for concentrating the cathode rays into a fine beam. The electron-optical system includes an element for controlling the current in the beam. The many forms of electron guns may be divided, broadly, into two classes; namely, those having purely electrostatic lenses, and those in which magnetic and electrostatic lenses are combined. Choice of the class of gun for a given cathode-ray tube depends upon the particular use intended and upon the operating conditions.

Most guns have a two-lens electron-optical system. The first lens is very close to the cathode and acts in such a way as to force the emitted

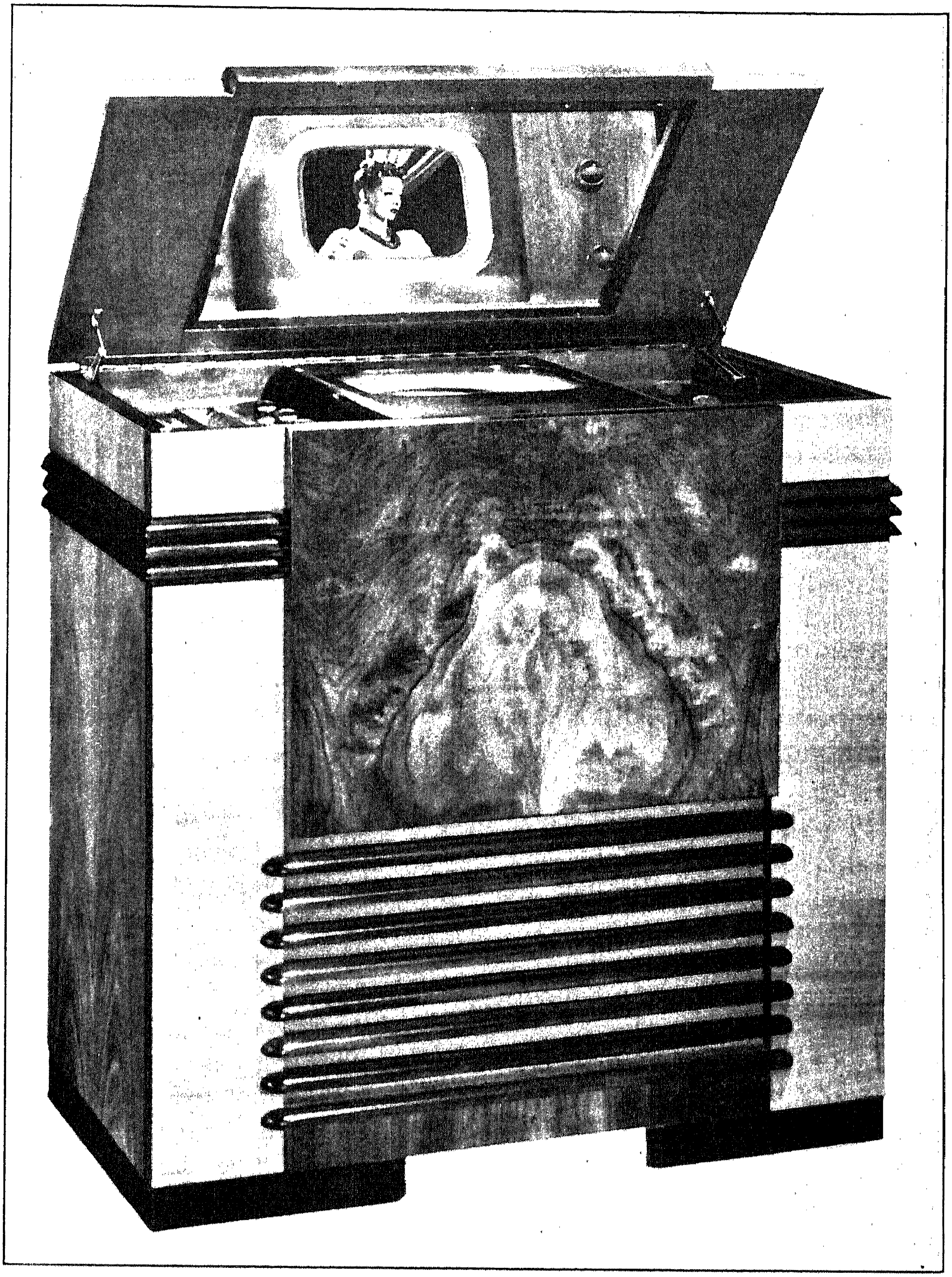


Fig. 9.20.—Console Television Receiver RCA-TRK12.

electrons into a narrow bundle, usually called the "crossover." The control grid, usually, is part of the first lens, governing the flow of electrons into the crossover by regulating the field in the immediate vicinity of the cathode. The crossover, which is the narrowest region of the beam within the gun, is imaged on the fluorescent screen by the second lens. This second lens may be either magnetic or electrostatic, depending upon the class of gun. The theory and design of an efficient electron gun are far from simple, and their importance warrants a more complete discussion, which will be given in Chapter 13.

The luminescent screen on which the picture is reproduced is made by coating the inside face of the tube with a fluorescent material in powder form. The nature and synthesis of the fluorescent material itself have been discussed in Chapter 2; the special problems involved in applying these materials to the glass wall will be dealt with in Chapter 12. Needless to say, the layer of fluorescent material must be uniform, free from contamination, and of the correct thickness.

The present-day screens give a black-and-white picture. This has been possible only within the last few years. The earlier television workers had to content themselves with the green and black reproduction characteristic of natural willemite. Fortunately for the development of the art, this naturally occurring material has a fairly high efficiency in converting electrical power into light; otherwise the acceptance of the electronic system would have been greatly delayed.

The size of the tubes commonly in use in home receivers permits screens ranging from 5 to 12 inches in diameter. The picture on the larger screen is 7½ by 10 inches. In the smaller, table model receivers the tubes are mounted horizontally, and the screen is viewed directly, as is illustrated in Fig. 9.19. A great many console receivers are so arranged that the tube rests in a vertical position with the screen upward, and the image is seen in a large mirror mounted in the cover of the set. When closed this cover not only protects the mirror and safety plate over the tube from mechanical injury but also keeps it free from dirt. In operating position the cover, to some extent, forms a hood which shields the tube from the overhead room lights. The brightness of the image reproduced on the screen ranges from 10 to 18 foot-lamberts so that the picture can be comfortably viewed in a normally lighted room. Fig. 9.20 illustrates a typical console receiver equipped for both picture and sound.

The physical size of the picture which can be obtained directly on the cathode-ray tube is limited by practical considerations of tube size. For the present, at least, the construction of sealed-off tubes having screens larger than 15 inches in diameter, although not impossible, is not very

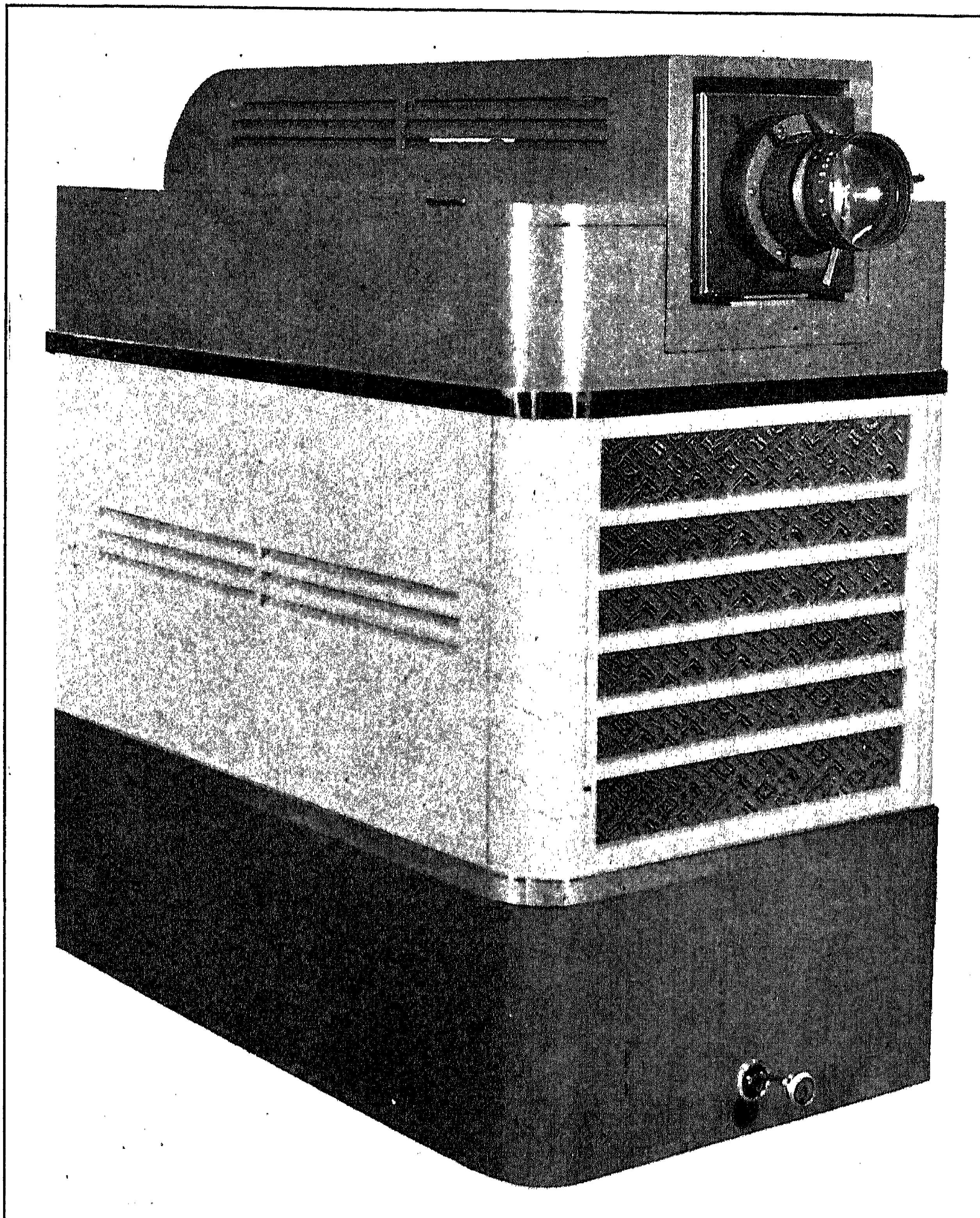


FIG. 9.21.—Projection Type Television Receiver (RCA Manufacturing Co.).

feasible. Much larger demountable tubes for direct viewing, however, are practical.

As has been pointed out, the cathode-ray tube can be used for projection purposes. The picture required in this case is much brighter owing to the relatively low efficiency of the ordinary optical system. The tube, in general, is smaller, as this facilitates the design of the projection lenses, although this small size puts severe demands on the electron gun.

A complete projection outfit, capable of forming a picture on a 5 by 7 foot screen, is shown in Fig. 9.21. This equipment uses an electrostatically focused projection Kinescope, operating at 20,000 volts. The image on the screen of this tube is 2¼ by 3 inches in size. When viewed on an ordinary directional moving-picture screen, the image is bright enough so that it can be watched in complete comfort; in fact, more incidental light in the room can be tolerated than is usually found in a moving-picture theater. Equally good results are obtained with magnetically focused projection tubes.

REFERENCES

1. J. C. Wilson, "Television Engineering," Pitman, London, 1937.
2. F. Schroeter (editor), "Fernsehen," Julius Springer, Berlin, 1937.
3. George Newnes, Ltd., "Television Today," London, 1936.
4. L. M. Myers, "Television Optics," Pitman, London, 1936.
5. I. G. Maloff and D. W. Epstein, "Electron Optics in Television," McGraw-Hill, New York, 1938.
6. H. W. Lee, "The Scophony Television Receiver," *Nature*, Vol. 142, pp. 59–62, July 9, 1938.
7. V. K. Zworykin, "Description of an Experimental Television System and the Kinescope," *Proc. I. R. E.*, Vol. 21, pp. 1655–1673, December, 1933.
8. Max Knoll, "Electron Optics in Television," *Z. tech. Physik*, Vol. 17, pp. 604–617, 1936.
9. R. R. Law, "High Current Electron Gun for Projection Kinescope," *Proc. I. R. E.*, Vol. 25, pp. 954–976, August, 1937.
10. V. K. Zworykin and W. Painter, "Projection Kinescope," *Proc. I. R. E.*, Vol. 25, pp. 937–953, August, 1937.

PART III

Component Elements of an Electronic Television System

CHAPTER 10

THE ICONOSCOPE

The role of the Iconoscope as a pickup device for the conversion of a light image into picture signals has been introduced in Chapter 8. The tube was described as consisting of a photosensitive mosaic and an electron gun assembled in a highly evacuated glass envelope. The gun is an electron-optical system serving to produce a fine pencil of cathode rays which is made to scan the sensitized side of the mosaic by means of a suitable magnetic, or electrostatic, deflecting system. The mosaic in the normal Iconoscope is a very thin mica sheet covered on one side with a vast number of minute silver globules, photosensitized and insulated from one another, and coated on the other with a metal film known as the signal plate. This metal film is coupled, on the one hand, to the silver elements by capacity through the mica and, on the other, to the video amplifier through a signal lead sealed into the bulb.

The optical image is projected onto the silver-coated side of the mosaic. Each silver element, being photoemissive, accumulates charge by emitting photoelectrons. Thus, information contained in the optical image is stored on the mosaic in the form of a charge image. The scanning beam sweeping across the mosaic in a series of parallel lines releases the charge from each element in turn and brings it to equilibrium, ready to start charging again. The change in charge in each element induces a similar change in charge in the signal plate and, consequently, a current pulse in the signal lead. The train of electrical impulses so generated constitutes the picture signal.

10.1. Construction of the Iconoscope. The appearance of a typical Iconoscope can be seen from Fig. 10.1. The shape and size of the blank containing the Iconoscope elements are matters of expediency, and, though the design illustrated appears to have certain advantages, many alternative forms have been successfully used.

The main body of the tube must be large enough to house a mosaic of suitable size. It will become apparent as the discussion proceeds that the magnitude of the picture signal is a function of the mosaic area. Therefore, the mosaic should be as large as is permitted by practical considerations of the associated optical system. Both the scanning beam and the optical images must have unobstructed access to the entire surface of

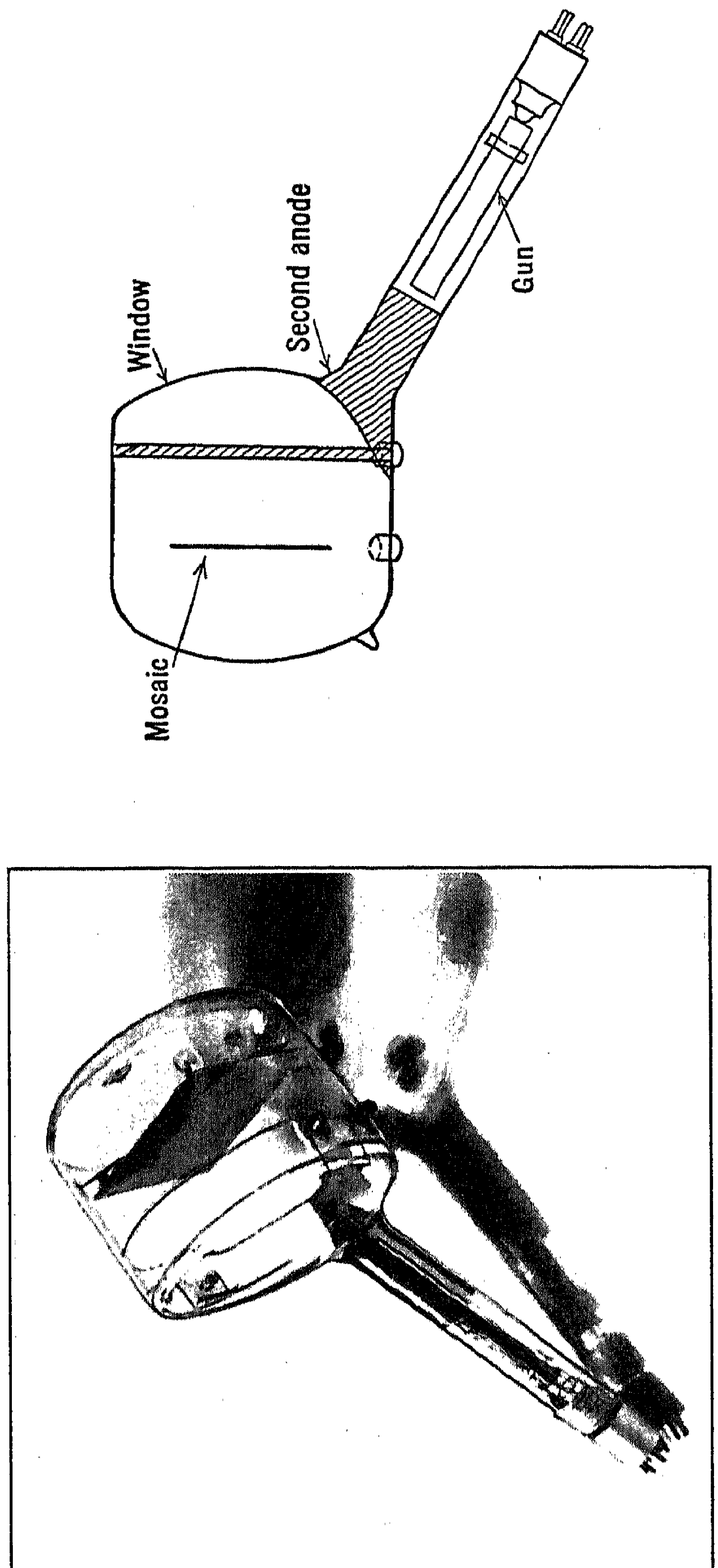


Fig. 10.1.—Photograph and Diagram of an Iconoscope.

the mosaic. Optical considerations require that the axis of the lens system which projects the picture coincide with the central normal from the mosaic surface. Hence, directly over the face of the mosaic there is an optical window free from imperfections that might distort the picture. The gun is placed at an angle to the normal through the center of the mosaic instead of coinciding with it as it would there interfere with the optical system. In the tube illustrated, the gun is mounted in the tubular arm forming the handle of the dipper-shaped blank. The optical window is the curved shell capping the blank on the gun side.

Two side tubes are required in the processing of the Iconoscope, one for the admission of caesium, the other for purposes of exhaust. The second tube is often located in the press at the end of the gun arm, for reasons which will be given later. The caesium tube may enter the blank at any convenient point, although to insure a uniform distribution of the alkali over the photosensitive surface a location behind the mosaic is advisable.

Several leads must be provided, of course, to carry the required potentials to the tube elements. Besides the wires in the gun press, two other leads are sealed into the walls of the tube, one to carry the picture signal, the other connecting the second anode. Additional leads may be necessary when special activation schedules, such as silver sensitization, are to be used.

The mosaic should be fastened rigidly in the blank in such a way that it will not warp or become misaligned during the processing. The mounting is purely a structural problem which admits of a number of solutions. In the tube shown in Fig. 10.1 the mosaic is supported on the metal cross member that can be seen in the illustration.

The gun structure shown in Fig. 10.2 is a characteristic example of the type used in the Iconoscope. It consists of a thermionic cathode, a control grid, and an insulating spacer, all mounted on a six-wire ring press, together with the first and second anode cylinders. The general method of assembly and the approximate dimensions of this gun can be seen from the figure. A small, closed, nickel cylinder serves as thermionic cathode and is indirectly heated by an internal tungsten heater. The cathode material is the usual barium and strontium oxide and coats the closed end of the cathode cylinder. The gun elements, all of which are cylindrical and mounted coaxially, can be assembled on an appropriate jig and spot-welded in place.

Essentially, the gun is a two-lens electron-optical system, one lens being located close to the cathode, the other between the first and second anode. The first lens is a multi-element system consisting of the cathode,

control grid, and first anode. As is evident from its name, the control grid serves to govern the current in the scanning beam. An accelerating grid, when employed, minimizes the effect of changes in the first-anode voltage on the beam current so that the focus of the gun may be varied without altering the current to the mosaic. Between the first and second anode a second lens is formed which serves to image the narrow section of the beam near the cathode, known as the crossover, onto the mosaic.

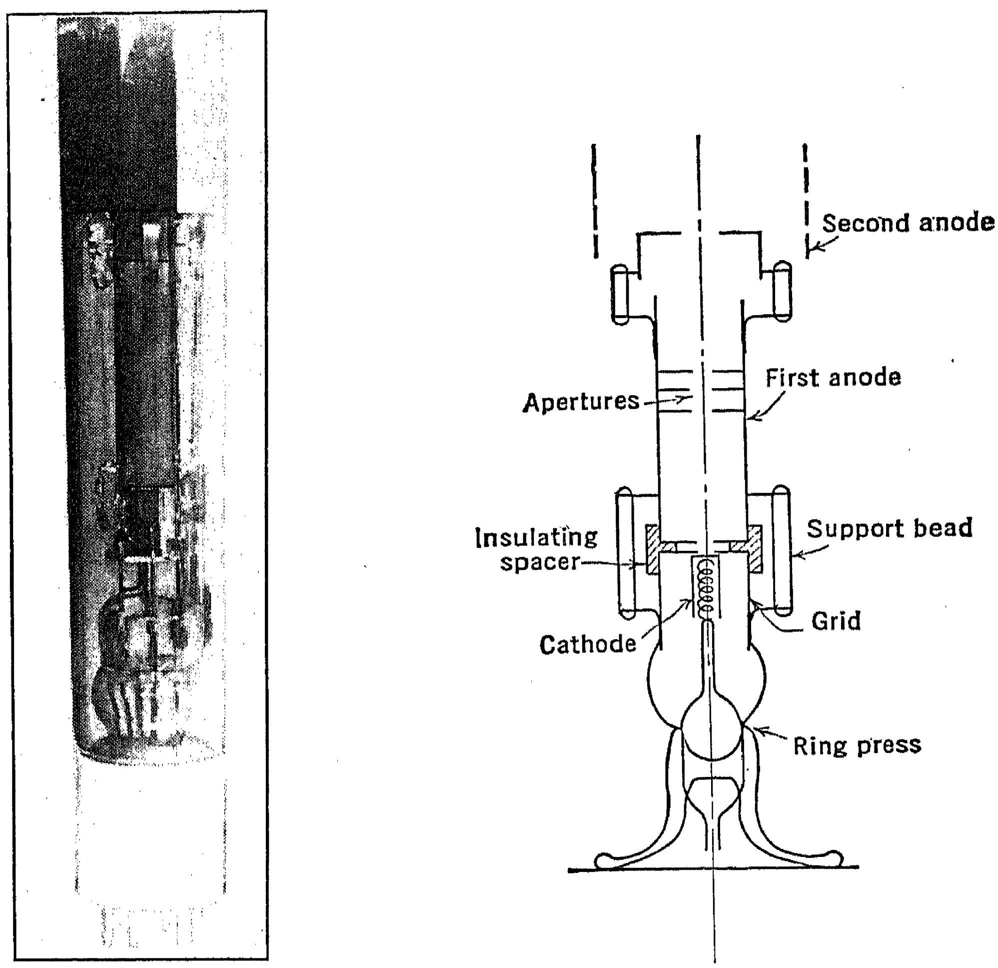


FIG. 10.2.—Iconoscope Gun.

Three apertures are mounted in the first anode cylinder in such a way as to form the aperture stop for the second lens and to mask off the secondary electrons emitted by the first anode itself. The second anode is divided into two parts. The first is a cupped washer attached to the main gun assembly at the end of the first anode; the rest is in the form of a metallized coating on the inside of the gun arm and extending into the main body of the tube. Since, as will be described later, the second anode also serves to collect electrons emitted by the mosaic, it also surrounds the latter usually in the form of a narrow ring. A more detailed account of the electron gun will be found in Chapter 13.

10.2. The Mosaic. The mosaic, as has been explained, consists of a dielectric sheet coated with a surface having high photosensitivity and low transverse conductance on one side and with a conducting film on the other. There are a great many ways of making such a mosaic, both as to the choice of dielectric and method of forming the sensitive layer.

The natural properties of mica make it a very suitable dielectric for the purpose. As a base for the mosaic of a normal iconoscope the mica should be cleaved into a thin sheet, the thickness being between 1 and 3 mils, depending upon the characteristics desired for the tube. In the subsequent theoretical considerations, a mosaic area of 100 cm^2 has been assumed. This requires a rectangle of approximately $3\frac{1}{2}$ by $4\frac{3}{4}$ inches. The sheet must be clean, preferably freshly cleaved, devoid of foreign inclusions, and free from steps due to fractured cleavage planes.

Since the cleavage operation is an important one, it is worth outlining an experimental technique for obtaining suitable sheets. A clear sheet of mica slightly larger than is required for the finished mosaic is selected. Separation of the sheets is started by gently tapping one corner. A strip of stiff paper is then inserted in the separation and worked around the edges of the mica, gradually moving toward the center. This process may sometimes be facilitated by immersing the mica in water (using a thin, celluloid spatula instead of the paper strip). Freshly cleaved mica, of course, requires no cleaning, and is ready for the next step in the preparation of the mosaic, namely, the formation of the silver globules.

The silver elements may be formed in a number of different ways. For example, a silver film may be evaporated onto the mica in the form of a continuous film and then, by heating, broken up into minute droplets; a film only a few atoms thick, so thin that it is not conducting, may be deposited on the mosaic after the tube has been evacuated; or silver oxide in the form of a powder may be spread over the mica and reduced to silver globules. Since the last-mentioned method lends itself well to experimental technique, it will be described in greater detail.

Silver oxide which has been ground in a ball mill until it is reduced to a fine flour is used to form the elements. A thin layer of the oxide is spread over the sheet by allowing the powder to settle through air. Without disturbing the coating the mica is heated rapidly to a temperature above the reduction point of the silver oxide. The rapid heating effects a reduction of the oxide without the individual droplets coalescing into a conducting film. It is often advantageous to lay down the moasic surface in several thin layers formed in succession, rather than completing it in one operation.

After the silver surface has been formed the back of the mosaic is metallized to make the signal plate. This order of building the mosaic

is usually preferable, particularly for experimental tubes, because it permits the use of transmission measurements to determine the coverage of the silver layer. The metallizing can be done by evaporation, sputtering, or with one of the many available chemical metallizing preparations.

In order to prevent the mosaic from buckling or warping, it is backed by a heavy sheet of mica to which it can be attached by means of a suitable strip metal frame. The backing sheet also facilitates making electrical contact to the signal plate.

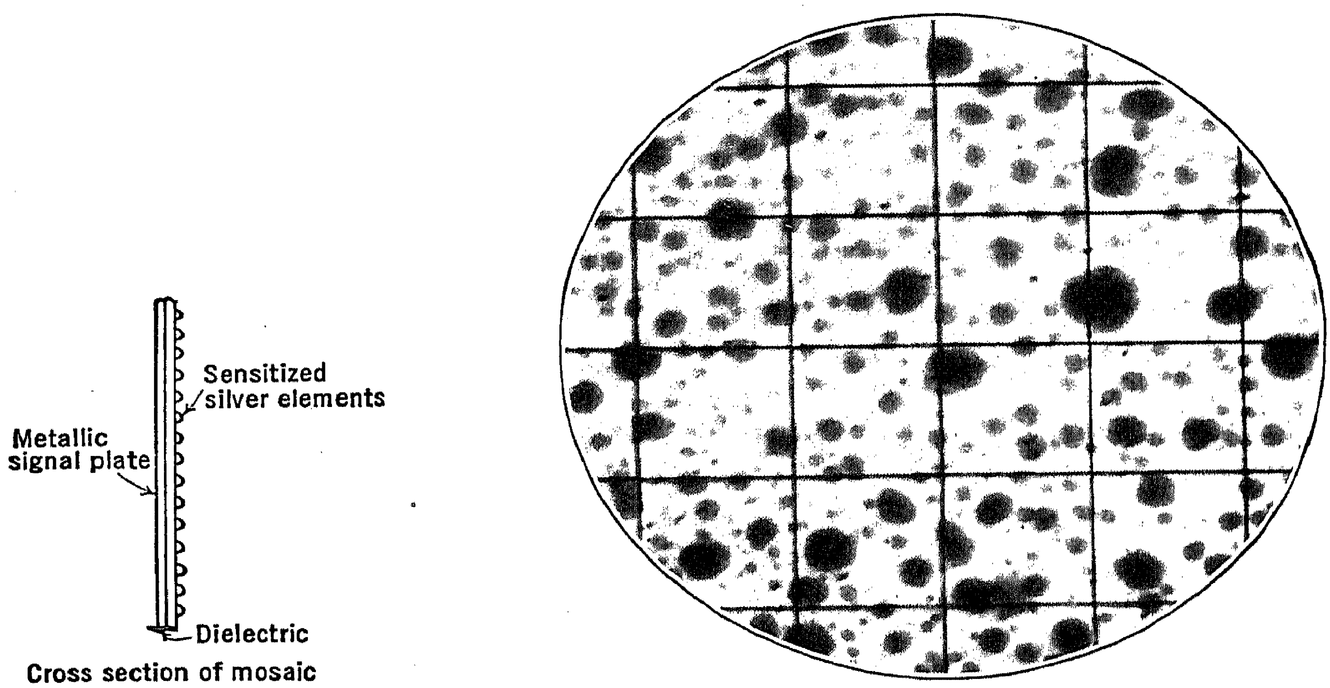


Fig. 10.3.—Structure of Iconoscope Mosaic.

Since the Iconoscope is essentially a high-impedance device, the friction contact between the signal plate and a metal strip threaded through slits in the backing sheet will be ample.

Before taking up the exhaust and activation procedure, other types of mosaic insulation should be considered. Various dielectrics such as enamels, metal oxide films, and thin powder layers may be substituted for cleavage sheets of mica.

An enamel to be used as dielectric must have a high specific resistance and be free from inhomogeneities. It is usually applied to a metal sheet, e.g., nickel, which serves as the signal plate. The coefficient of thermal expansion of the coating must match that of the underlying metal.

Metal oxide films can be formed by chemically oxidizing the metal of the signal plate. Aluminum oxide on an aluminum signal plate can easily be thus prepared. Other oxide dielectrics, such as barium borate, magnesium oxide, or silica, may be evaporated onto the signal plate.

Powder insulated mosaics have been found to be very satisfactory, even surpassing mica in some respects. Flowing, or spraying, a liquid suspension of the powder over the signal plate is fairly satisfactory, but other and more complicated techniques lead to much better results. Magnesium oxide, kaolin, and colloidal quartz are all suitable as dielectrics for powder mosaics.

10.3. Exhaust and Activation Schedule. The assembled Iconoscope is sealed to the vacuum system by the exhaust tube in the gun press. It should be noted that, though this location of the exhaust vent is not essential, it reduces the possibility of the mosaic's being contaminated by gases given off by the gun during activation. Caesium required for activation is admitted through the side tube already mentioned. Convenient sources of caesium have been described in Chapter 5.

The exhaust requirements for the Iconoscope are severe. While the final pressure may be as high as 10^{-5} mm of mercury in an operative tube, the slightest trace of certain impurities is found to be absolutely fatal. The evacuation schedule for the most part follows standard practice. The tube is pumped to a low pressure at room temperature, and then heated to as high a temperature as the blank can safely withstand. During the initial exhaust and the first fifteen to thirty minutes of bake, a freezing-out trap normally is not used. This avoids collecting at the mouth of the trap large amounts of condensable vapors given off by the glass and metal parts. For the rest of the exhaust, the trap is used. The bake is continued until the pressure in the tube has dropped to 4 or 5×10^{-5} mm of mercury, after which the Iconoscope is allowed to cool slowly to room temperature. In order to drive out gases occluded in the metal parts of the gun, these parts are brought to a dull red heat with a high-frequency bombarder.

The activation of the mosaic follows closely that used for the hard caesium phototube, with this difference: in the phototube the only consideration determining the amount of caesium is the sensitivity, whereas in the mosaic both sensitivity and conductivity across the surface are dependent upon the amount of caesium. The process of sensitization consists of oxidizing the silver surface, admitting caesium, and baking to promote a reaction between the silver oxide and the alkali.

Oxidation of the silver elements on the mosaic can easily be effected by the procedure given in Chapter 1. Oxygen is admitted into the system to a pressure of about 1 mm, and a glow discharge is formed over the surface of the mosaic. The glow discharge can be produced with the aid of a high-voltage, high-frequency oscillator grounded on one side, the other side being connected to a movable electrode placed outside the walls of the tube. A very uniform oxidation can be obtained by moving

the electrode judiciously over the bulb. The thickness of the oxide layer can be gauged by the interference colors produced by the film on the metal. The colors appear in the usual order for subtractive interference colors as the thickness is increased, namely:

Yellow.	Second yellow.
Red.	Second red.
Blue.	Second green.
Blue-green.	etc.

As in phototube practice, the first blue will produce a very satisfactory photosensitive mosaic.

Care should be taken to avoid a violent discharge during oxidation as this will cause silver from the globules to be sputtered over the mica, resulting in loss of both signal and resolution due to leakage between the elements. After the oxidation the excess oxygen is pumped out of the system. A short bake at 200°C will expedite the complete removal of the oxygen, but is not essential.

The thermionic cathode in the electron gun employs an emitting surface of barium and strontium oxide. The preparation and activation of this type of cathode were described in Chapter 1. In activating a cathode prepared in this way, a large amount of gas is released during the conversion of the carbonate to the oxide. As this gas may contaminate a sensitized mosaic, it is advisable to activate the cathode before proceeding with the caesiation. During the activation of the cathode the grid structure should be kept at a dull red heat to avoid the formation of an oxide-bearing film.

The final step in processing the Iconoscope is the caesiation of the mosaic. A quantity of caesium, slightly less than that required for complete activation, is driven into the Iconoscope. In order to promote a reaction between the silver oxide and the caesium, the tube is baked at 200°C for a few minutes. After it has again become cool the sensitivity is measured ballistically, by methods described in the next section. The final sensitivity should lie between 7 and 10 microamperes per lumen. By comparing the first sensitivity reading with this figure, the additional caesium necessary can be estimated. After this amount has been admitted the tube is again baked. Care should be taken to avoid an excess of caesium, as this will result not only in poor photosensitivity but also in loss of both resolution and storage due to leakage. If an excess has accidentally been admitted, its harmful effects can sometimes be reduced by a prolonged bake at 200°C. The presence of a tin oxide or lead oxide caesium getter in the tube will also reduce the harmful effect of too much caesium.

The simple sensitization schedule described is well suited to meet experimental laboratory needs. In the commercial production of Iconoscopes a much more exacting procedure is employed. Various rather complicated techniques, such as sensitization with silver and argon or hydrogen glow discharges, have been developed which not only increase the sensitivity of the tube but also improve the color response. These methods lead to more than double the sensitivity resulting from the simple procedure outlined.

10.4. Performance Tests. The ultimate test of the merit of an Iconoscope is, of course, its performance in the television system. There are, however, a series of measurements which can be made during its construction and activation, before it is actually used, and which will give an indication of quality and thus greatly facilitate producing a good tube.

The Iconoscope must be tested for its behavior with regard to the following factors:

(*a*) Photosensitivity.
(*b*) Leakage over mosaic.
(*c*) Characteristics of gun.
(*d*) Perfection of mosaic.
(*e*) Resolution.
(*f*) Sensitivity.
(*g*) Color response.
(*h*) Spurious signal.
(*i*) Secondary emission from gun.
(*j*) Miscellaneous defects.

Of these tests, the first three are made during the activation of the Iconoscope, the remaining after its completion.

Test (*a*) for photosensitivity is extremely important and should be made at regular intervals during the final stages of the caesiation of the tube, so that it can serve as a guide to the treatment required.

Since the mosaic is an insulator, transient measurements are necessary to determine the photoemissivity of its surface. The ballistic method described below is found convenient and is often used.

The method consists of applying a potential between the signal plate and second anode, and then allowing a flash of light of known intensity and duration to fall on the mosaic. The charge released as a result of the light impulse is measured with a ballistic galvanometer. The set-up for this measurement is shown in Fig. 10.4. The light source projects a spot of light about 2 inches in diameter onto the mosaic. The duration of the light impulse is controlled by the shutter B; an ordinary camera shutter will serve the purpose admirably. The quantity of light found convenient is of the order of 1/100 lumen-second, that is, for example, ½ lumen for 1/50 second.

Since the emission from an average mosaic lies between 7 and 15 microamperes per lumen, the charge released will be between 0.07 and 0.15 microcoulomb. Furthermore, since the capacity of the mosaic to its

signal plate is of the order of 100 micro-microfarads per square centimeter, the rise in potential will be of the order of 60 volts. The voltage, V, applied to the second anode must be well in excess of this value to insure that the photoemission is at all times saturated. In practice, a potential of 200 to 400 volts is used.

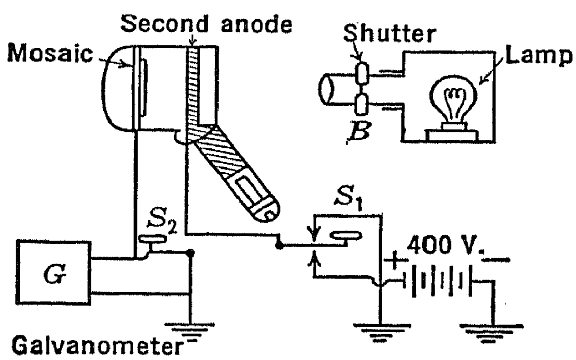


Fig. 10.4.—Ballistic Sensitivity Test.

When the test is started, the galvanometer G is short-circuited through switch S_2 and the second anode grounded by means of S_1. The whole Iconoscope is then illuminated with an intense light so that photoemission from the second anode and other metal parts will bring the mosaic to approximately ground potential. Next, the bulb is so covered that no light can reach the mosaic except through the shutter from the lamp. Switch S_1 is moved to the other contact, while S_2 still shunts the galvanometer to prevent the capacity surge from affecting the instrument. The galvanometer switch S_2 is then opened and the light flashed. From the charge indicated by the galvanometer deflection and the quantity of light the photosensitivity can be calculated.

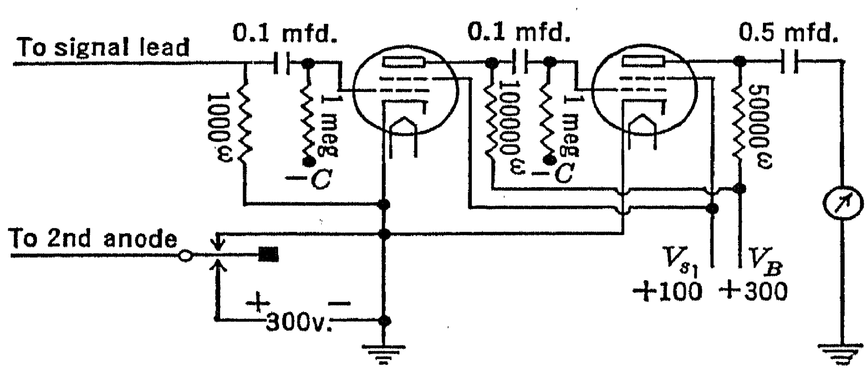


Fig. 10.5.—Vacuum-Tube Meter for Sensitivity Test.

An alternative, very similar to the above method, is to illuminate the mosaic continuously, and to apply the second anode potential for a short known period of time. In general, this is found to be less satisfactory than the former procedure.

The ballistic galvanometer for these tests may be replaced by a vacuum-tube meter. A circuit suitable for the purpose is shown in Fig. 10.5.

Next in importance is the measurement of leakage over the mosaic. This can be made as follows. With the test equipment just described, and the potential applied to the second anode, sufficient light from the source

is allowed to fall on the mosaic to cause the illuminated portion of the mosaic to reach second anode potential. The light required for this will be of the order of 1/10 lumen-second. The mosaic is then left in darkness for 5 seconds. The deflection produced by a 1/100 lumen-second impulse of light then applied to the mosaic will indicate the amount of charge which has leaked away from the charged area into uncharged regions. With 200 volts applied to the second anode the deflection thus observed should be small compared with that obtained in the normal sensitivity test. It should be noted that the time of exposure required to saturate the area of the mosaic illuminated will depend upon the sensitivity. If leakage tests are made during the early stages of activation, when the sensitivity is low, the exposure time must be lengthened accordingly. However, the exposure time should be no longer than necessary as stray light and leakage may cause the unexposed portions of the mosaic to become saturated, thus vitiating the leakage measurement. Because the requirements of low leakage and high photosensitivity must both be fulfilled in a satisfactory Iconoscope, the importance of these tests cannot be overemphasized.

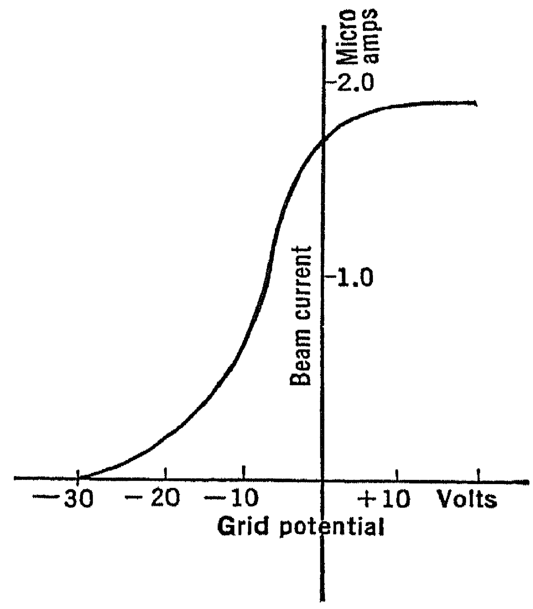


Fig. 10.6.—Current Characteristic of a Typical Iconoscope Gun.

Test (*c*), of the gun performance, can be simply performed by applying the normal voltages to the gun and measuring the current to the second anode as a function of grid bias. In carrying out this test the beam should be deflected off the mosaic with a small permanent magnet to avoid burning a spot on the sensitized surface. A typical characteristic curve of the gun is shown in Fig. 10.6.

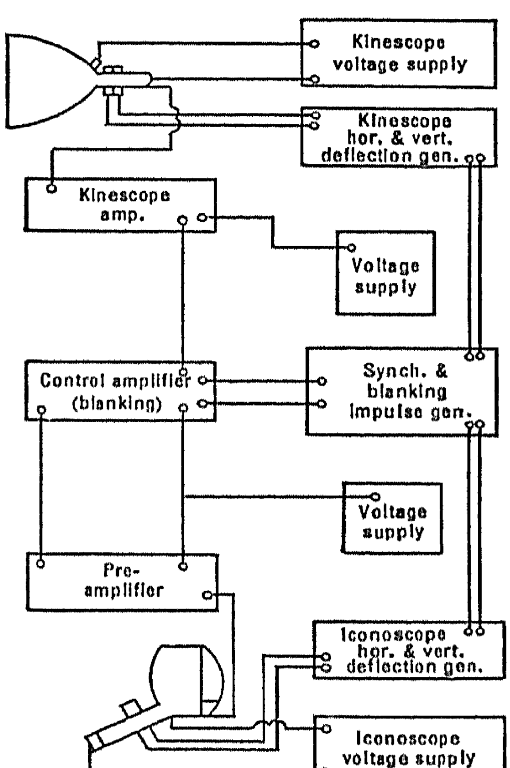


Fig. 10.7.—Block Diagram of an Iconoscope Test Set.

The remaining tests are performed after the Iconoscope has been completed and sealed off from the vacuum system. These tests require equipment consisting of a video amplifier with an available

gain of at least 8000, voltage supplies suitable for providing the potentials needed by the Iconoscope, horizontal and vertical sawtooth generators, and a Kinescope for viewing the reproduced picture. Descriptions of these various units will be found in the succeeding chapters of this book.

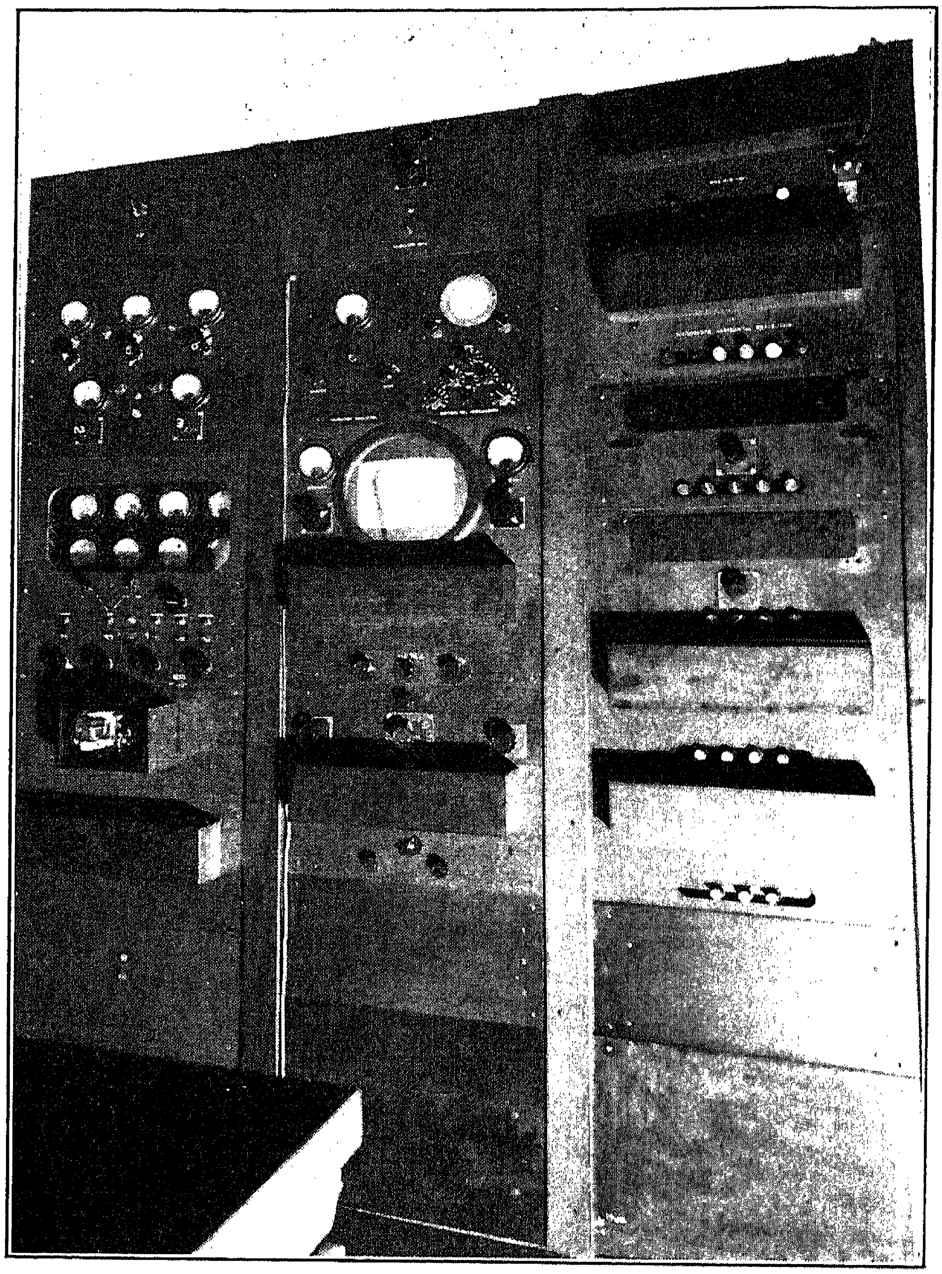


FIG. 10.8.—View of a Typical Iconoscope Test Set.

A simplified diagram of the complete test outfit is shown in Fig. 10.7; Fig. 10.8 is a photograph of a typical Iconoscope test outfit.

The completed Iconoscope is connected into the test set and the voltages adjusted to the values required for correct operation. With the tube in darkness the beam is allowed to scan the mosaic. The resulting picture on the Kinescope should show the mosaic as a uniform clean surface. Small bright or dark spots are usually due to dust or dirt particles on the

mosaic, or to pinholes in the signal plate. Large, irregular patches may be due to dirt, to non-uniform distribution of caesium, or to lack of uniformity in oxidation. Sharp, straight-edged areas are due either to splits across cleavage planes of the mica or to scratches on the signal plate.

A uniform light is then projected over the surface of the mosaic to determine the uniformity of photoemission.

To test the definition of the tube, a resolution pattern, such as is reproduced in Fig. 10.9, is projected onto the mosaic. A good experimental Iconoscope should resolve the pattern down to at least 600 to 800 lines. Since often the video amplifier response is limited to 3 or 4 megacycles,

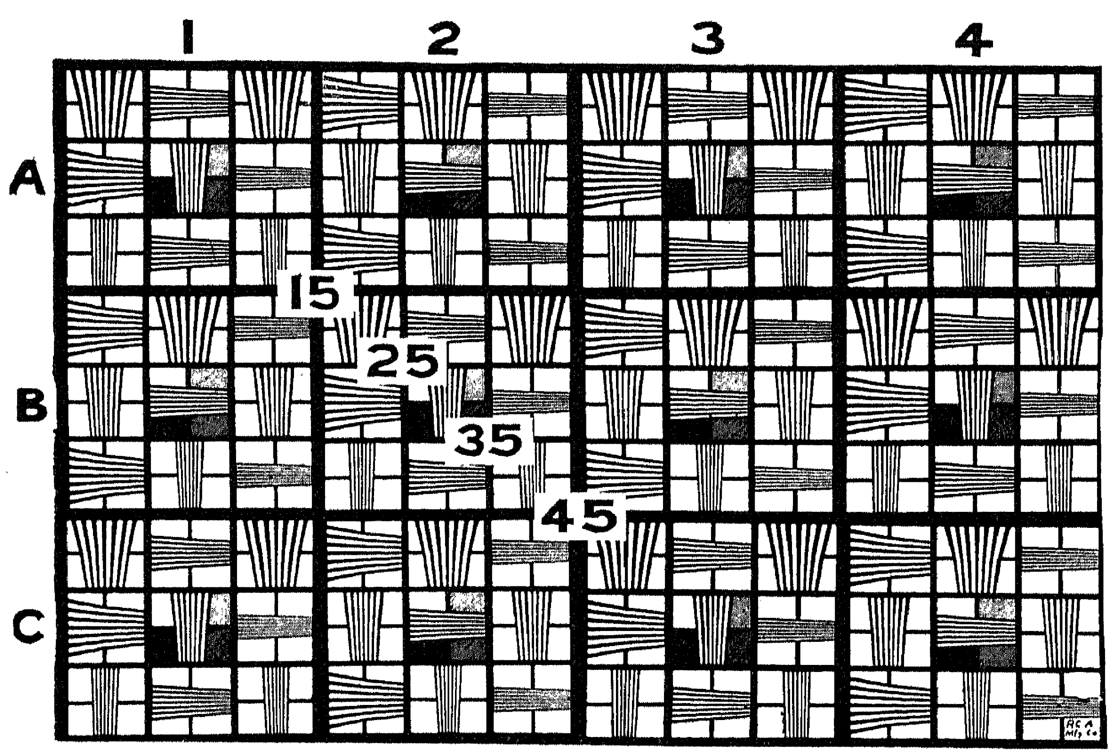


FIG. 10.9.—Resolution Pattern.

the resolution of the picture seen on the Kinescope under normal conditions may be limited by the system. However, if the horizontal scanning is reduced to cover only a small fraction of the mosaic while keeping the Kinescope deflection normal, an enlarged image of the pattern can be seen on the screen. The resolution of this enlarged section will be limited by the Iconoscope only.

It is important that the spot produced by the gun be round, as any ellipticity due to misalignment will cause unequal resolution in various directions. The Iconoscope deflection being contracted both vertically and horizontally, the resolution obtained along different axes will be a measure of the regularity of the spot.

For a routine estimate of the sensitivity, a picture of known quality,

such as the resolution pattern just illustrated, is projected onto the mosaic and the image is observed as the light in the projected picture is reduced. The amount of light required to produce a picture just visible over the amplifier noise is a measure of the sensitivity. This illumination can be compared with that required to produce a similar picture in an Iconoscope of known performance.

A quantitative measurement of the sensitivity is much more difficult. It is usually made by measuring the signal output, when a small area of the mosaic—for example, a square covering an area equal to 1/100 the total area of the mosaic—is illuminated with a known light intensity. A plot is then made of signal output against light intensity. Typical curves will be found in section 10.11.

The relative color response can be determined by projecting a spectrum of known light distribution onto the mosaic and measuring the signal output for the various wavelengths. A color chart of the type used for photographic measurements imaged on the mosaic affords a very convenient way of estimating the color response.

When the mosaic is scanned, even in darkness, a spurious signal in the form of shading will be observed on the viewing screen. This is variously known as "black spot," "tilt and bend," "shading," etc., and its cause will be discussed in a succeeding section. The magnitude of the signal increases rapidly with beam current. In a good tube the spurious signal should be small and fairly regular. Electrical compensation of this signal will be taken up in a later chapter.

Appreciable secondary emission from the electron gun is rather rare and is always due to some fault in the gun construction. The symptoms are a haze which resembles a partially transparent curtain over the picture. If the amplitude of the deflection is greatly reduced, this haze will be seen to take the form of a very blurred image of the mosaic. When this defect is present it can be overcome by operating the gun under conditions of inverse focus. This is done by applying a higher potential to the first anode than to the second anode. For example, a three-element gun which focuses at a ratio of 1:4 between first and second anode will be found to focus also with a ratio around 3:1. However, this mode of operation is not very satisfactory, from the standpoint either of the resolution obtainable or of the life of the tube.

The final test consists of examining the performance of the tube under as wide a variety of operating conditions as possible. This test should reveal whether the tube oscillates under high beam current, whether the scanning beam strikes the neck of the gun, and any other defects which might lessen its usefulness.

10.5. Theory of Operation—Characteristics of the Mosaic. In preceding chapters a very much simplified picture of the operation of the Iconoscope was given. The actual details of operation are extremely complicated, and as interesting as they are complex. Quantitative measurements of the various factors involved in the operation are difficult, and in general require the introduction of measuring electrodes and probes, which alter to some degree the conditions existing in the normal tube. The extent of these changes cannot readily be estimated. However, it is possible to derive a working theory yielding quantitative results which are in fair agreement with the measured performance of the Iconoscope.

The surface of the mosaic, as has been pointed out, consists of a vast array of minute photosensitive elements. The size of these elements depends upon the way in which the mosaic was formed and sensitized, and may range from droplets 0.0005 cm in diameter down to submicroscopic particles. In any case, these elements are very small compared to the diameter of the scanning spot, so that a large number of these particles are under the beam at any one time. Thus, it is evident that, from the standpoint of the mosaic, the concept of a picture element has no real significance, but is only a convenient fiction for simplifying the description of the operation of the Iconoscope. In what follows, the mosaic will be treated as a two-dimensional continuum capable of photoelectric and secondary emission but with zero transverse conductance.

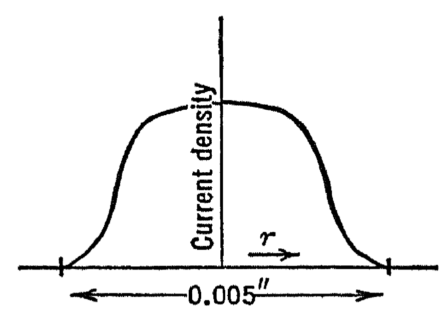


FIG. 10.10.—Current Distribution in Scanning Spot.

The capacity of the sensitized surface to the signal plate depends upon the thickness of the insulating layer and upon the dielectric material. In general, it may be said to range between 50 and 300 micro-microfarads per square centimeter. The operation of the ordinary Iconoscope is not very critical to the value used. For convenience, a value of 100 micro-microfarads per square centimeter will be assumed.

The secondary-emission ratio of the mosaic surface also may vary over a wide range. It has been measured to be as low as 2 for certain types of mosaics, to 8 or 9 for others. A fair average value to take as a basis for discussing the behavior of the tube is 4. The photoelectric sensitivity of a good silver-sensitized mosaic is in the neighborhood of 15 microamperes per lumen.

The electron scanning beam at the mosaic has a diameter of 0.01 to 0.02 cm and a current of 0.1 to 0.25 microampere. The current distribution in this beam is shown in Fig. 10.10.

For purposes of discussion, the spot will be considered as a square 0.02 cm on a side and having a uniform current density of 5×10^{-4} amp/cm². As it moves across the mosaic it has a linear velocity of 1.6×10^5 cm/sec. This assumes a picture period of 1/30 second and a 441-line picture. For simplicity it will be further assumed that the scanning is not interlaced. The effect of interlaced scanning will be discussed somewhat later. The dimensions of the mosaic commonly used are 3½ by 4¾ inches, or approximately 9×12 cm—an area of slightly more than 100 cm². On the basis of the average capacity per unit area given above, the total capacity of the sensitized surface of the mosaic to the signal plate is about 10,000 micro-microfarads. In this connection it may be mentioned that the capacity between signal plate and second anode is in the neighborhood of 6 micro-microfarads.

10.6. Potential Distribution on the Mosaic. Contrary to what might be expected, the surface of the mosaic when scanned in darkness does not assume a uniform potential. Measurements made both on a caesiated silver element set into the surface of a mosaic and connected by a lead through the wall of the tube, and on tubes having a divided signal plate, show that directly under the scanning beam the mosaic assumes its most positive potential, and that it is at its most negative potential directly in front of the scanning beam. Between these two points there is a continuous transition of potential. At the point of bombardment the surface assumes a voltage between +2 and +3 volts with respect to the second anode. This potential is, to some extent, dependent upon the physical state of the surface of the mosaic and upon the velocity of the bombarding electrons, but it is to the first approximation independent of the current in the scanning beam. On the other hand, the lower limit of the potential is primarily a function of the beam current and secondary-emission characteristics. At very high currents, that is, between 0.5 and 1.0 microampere, the lower potential will be about −1½ volts. As the current is decreased, the lower limit gradually approaches the potential under the beam. Fig. 10.11 is a map of the potential distribution over the mosaic for various beam currents.

The rather complicated potential distribution arises from the fact that two different types of potential equilibria are possible simultaneously at different points on the mosaic. These are: first, that of an insulated secondary-emissive area under electron bombardment; and second, the retarding potential necessary to prevent electrons scattered from the point under the bombarding beam from reaching the mosaic.

In order to account for the potential of the point directly under the bombarding beam, it is necessary to consider the secondary-emission phenomenon taking place at this area. Since the mosaic is surfaced with

caesiated silver which has been activated according to the schedule used for the Iconoscope, a saturated secondary-emission ratio of 4 can be assumed. When an insulated surface of this type is bombarded, more electrons will leave than arrive, causing it to become positive with respect

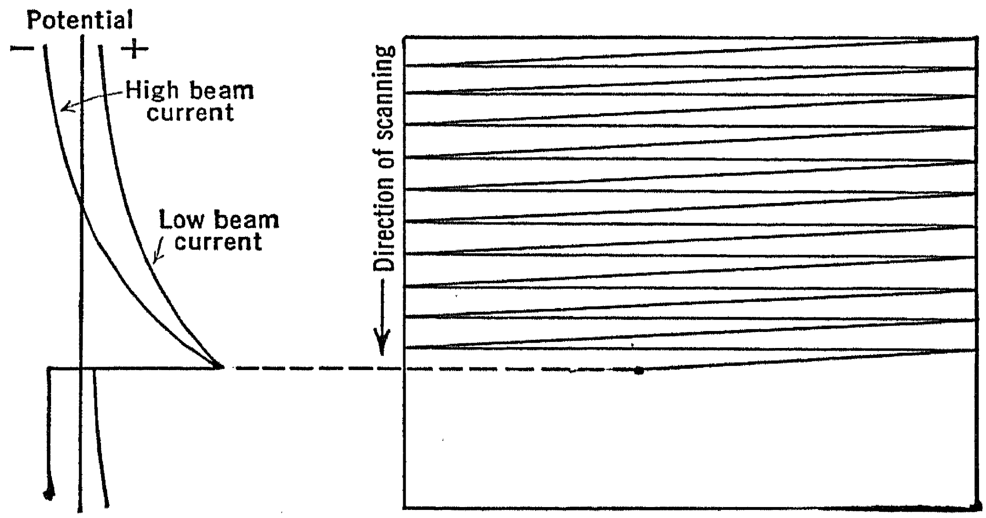


FIG. 10.11.—Distribution of Potential on Mosaic.

to the electrodes which collect the secondary electrons. As the potential becomes increasingly positive it more and more prevents secondary electrons having small initial velocities from reaching the collecting electrode. When the potential reaches such a value that the escaping current just equals the bombarding current, equilibrium is established and there is

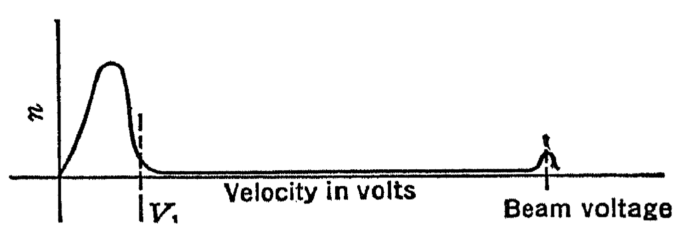


FIG. 10.12.—Velocity Distribution of Secondary Electrons from a Mosaic.

no further change in potential. To clarify this mechanism further, Fig. 10.12 is presented. The curve represents the velocity distribution of secondary electrons from a surface of the type under discussion. The abscissa of the curve gives the initial velocity in volts; the ordinate, n, multiplied by the velocity range dV, gives the number of electrons having velocities lying between any value V and $V + dV$ for each primary electron. The integral of this curve:

$$\int_0^\infty n\,dV = R$$

gives the secondary-emission ratio R. Again, if the emitting surface has a potential V_1 the ratio of secondary to primary current will be:

$$R' = \int_{V_1}^{\infty} n dV.$$

Equilibrium will be established when R' is equal to unity. The equilibrium potential for a surface of the type found in the mosaic lies, as has been stated, between 2 and 3 volts.

Owing to the positive potential of the point just considered, and to free electrons between the mosaic and second anode, which is the collecting electrode in this case, a certain fraction of the electrons is returned to points on the mosaic not under the scanning beam. This current, which reaches the mosaic in the form of a more or less uniform rain of redistribution electrons, causes the rest of the surface to become increasingly negative. When a region of surface has reached $-1\frac{1}{2}$ volts, the retarding potential is sufficient to prevent all or nearly all redistribution electrons from reaching this area. Thus, the limits of the potential variation over the surface are $+3$ volts and $-1\frac{1}{2}$ volts. If the beam current is small, the redistribution current will not be sufficiently great to cause any point on the mosaic to reach the lower equilibrium point in a frame period.

Thus, to summarize the cycle which occurs when the mosaic is scanned in darkness, it is found that an element of area under bombardment assumes a potential of $+3$ volts. As the scanning beam moves on, the element receives a small current in the form of redistributed electrons. This causes the area to become increasingly negative as the beam moves farther away. The decrease in potential continues until either the scanning beam returns to the element in question or until it reaches about $-1\frac{1}{2}$ volts.

10.7. The Mosaic under the Influence of a Light Image. The mechanism of the generation of the picture signal can be described on the basis of the phenomenon just discussed. Fundamentally, it is due to the fact that the potential of an illuminated element of area, just before being reached by the scanning beam, is more positive than that of one which has not been illuminated. Since the mosaic is driven to the same positive equilibrium regardless of its initial potential, charge will be released at a different rate when the beam is traversing the lighted region from that when scanning an unlighted area. The change in current reaching the second anode constitutes the picture signal.

The difference in potential due to light is, of course, caused by the storing up of the charge resulting from the photoemission of the illuminated areas. Under actual operating conditions this photoemission is less

than the redistribution current so that, relative to the second anode, every element of area becomes increasingly less positive with respect to the second anode. Those which are illuminated decrease in potential less rapidly than the remainder.

The photoemission from the mosaic is, in general, unsaturated, as will be evident when it is remembered that the potential of the mosaic is, for the most part, positive with respect to the second anode. This makes the calculation of the signal to be expected from a given amount of illumination very difficult because it means that the emission is a function of the potential of the area under consideration as well as of the incident light. A further complication results from the fact that it is entirely unnecessary for the photoelectrons to leave the mosaic in order that a picture signal be generated. In fact, under normal operating conditions, the fraction of photoelectrons from an illuminated area which actually leave the mosaic is insignificant compared to that which is returned to the mosaic in the form of redistribution electrons. Experimentally, this can be easily shown by interrupting the image projected onto the mosaic at a frequency which is some multiple of the frame frequency. If care is taken to insure that the second anode and signal plate are not photoemissive, almost no signal will be observed corresponding to the interruption frequency. This indicates that no photoelectric current is traversing the circuit consisting of mosaic, second anode, amplifier coupling impedance, and signal plate.

In view of the complicated nature of the unsaturated photoemission responsible for the picture signal, it will not be amiss to discuss it in greater detail. Considering first the extreme case of a mosaic in the absence of a scanning beam, if a small area of the mosaic is illuminated with an intensity L_1, this area will emit electrons which in part go to the second anode and in part return to the mosaic. Eventually this area becomes sufficiently positive to prevent the escape of further electrons. It is obvious that under these conditions the positive potential assumed by the area is dependent upon the initial velocities of the photoelectrons rather than upon the intensity of illumination. Therefore, if an intensity L_2—half that of L_1—had been used, the potential would have been approximately the same. If, however, two areas are illuminated at the same time, one with an intensity L_1, the other with L_2, these areas will not come to the same potential. This will be clear if the photoelectric saturation curve shown in Fig. 10.13 is examined and the behavior of the electrons considered. The photoemission curves are represented as straight lines, for convenience of discussion, rather than in their true form. The ordinate represents the actual number of electrons emitted by an elementary area of the mosaic at any potential, and not the net current,

which is the algebraic sum of the current leaving and the redistribution current arriving from other portions of the mosaic. When the two areas emit simultaneously the return current to every illuminated element of area will be approximately equal.* This is represented by the line ab. In order that the net current to each elementary area be zero, the two illuminated regions must take on the potentials E_1 and E_2, respectively. Similarly, if an entire image is projected onto an unscanned mosaic, a corresponding potential image will be established as an equilibrium condition.

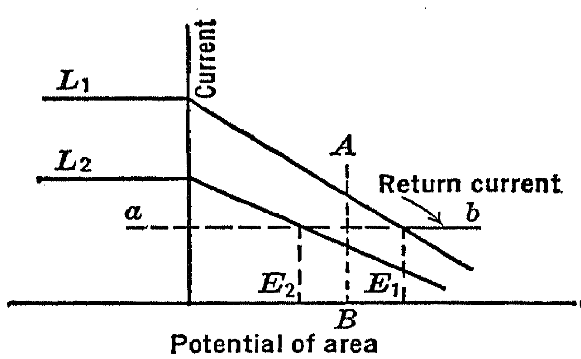


FIG. 10.13.—Current-Density Characteristics of Differently Illuminated Mosaic Elements.

Under operating conditions, when a relatively weak beam is used and an intense image is projected onto the mosaic, the mechanism of the formation of the charge image will be essentially that outlined above, except that the current represented by ab will include not only the returning photoelectrons but also the redistribution electrons from the beam. In this case the charge image is fully established in less than a picture period, and further exposure produces only a slight increase in picture signal. (This increase is due to line sensitivity, which will be discussed in detail in succeeding pages.) To avoid possible misunderstanding, it should be pointed out that this saturated mode of operation does not imply that the tube is operating as a non-storage system, since, even if the picture is fully established in a time corresponding to, say, 10 lines, the storage factor is still some 5000 times. As the illumination is decreased, eventually a point is reached where the charge image is no longer fully established in a frame time. Under these circumstances the illuminated area is receiving redistributed photoelectrons and beam secondary electrons, and during the cycle becomes increasingly negative with respect to the second anode; however, its potential is continuously rising with respect to an unilluminated area. Fig. 10.14a shows the time variation of the potential of an illuminated area and an unilluminated area with respect to the second anode, while b shows the potential of the same illuminated area referred to that of the unilluminated region.

From a more quantitative point of view, if an element of area not under the beam ds_a (Fig. 10.15) is considered it is found to receive the

* Strictly speaking the magnitude of the return current is a complex function of tube geometry, mosaic activation, distribution of illumination, etc.

following current components: ρ_p the current density due to photoelectrons leaving the area, ρ_{r1} redistribution electrons from the beam, and

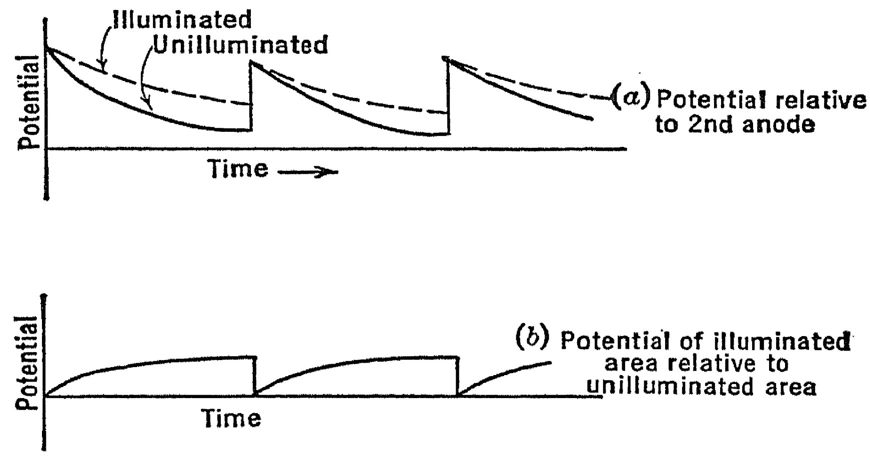


FIG. 10.14.—Variation of Potential with Time for Unilluminated and Illuminated Elements.

ρ_{r2} returned photoelectrons. The rate at which its potential is changing with respect to the second anode will be:

$$\frac{dV_a}{dt} = \frac{(\rho_p - \rho_{r1} - \rho_{r2})}{C_0 ds_a} ds_a,$$

where C_0 is the capacity per unit area. The corresponding rate for an element ds_b, in darkness, will be:

$$\frac{dV_b}{dt} = \frac{-(\rho_{r1} + \rho_{r2})}{C_0 ds_b} ds_b.$$

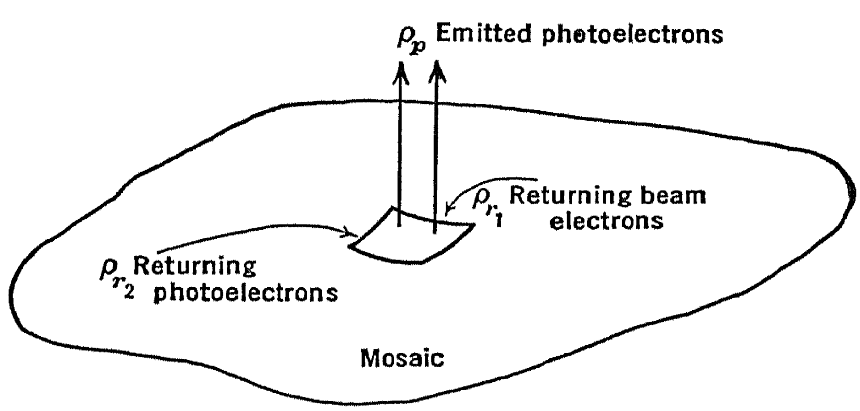


FIG. 10.15.—Current Components for an Element Not under the Beam.

The three current-density components in the above differential equations are functions of the potential with respect to the second anode and of the difference in potential between various portions of the mosaic. Actually, however, the variation of ρ_{r1} and ρ_{r2} with local variations in potential is relatively small. Assumptions could be made as to the func-

tions involved, and the differential equations could be solved. However, since little is known as to the nature of these functions, very little would be gained by doing this. It must suffice to say that, under conditions of low light levels, the average photoelectric current leaving a small, illuminated region, with the rest of the mosaic in darkness, is approximately 20 per cent of the saturated photoemission.

Summarizing the portion of the cycle discussed above, directly under the beam the mosaic assumes a potential V_2; examined farther and farther from the point under bombardment, the potential is found to decrease to some potential V_1 (x,y), which is not constant over the mosaic but is a function of the light distribution in the image and, therefore, a function of the coordinates, x,y, on the mosaic.

10.8. The Formation of the Video Signal. The final step in the cycle is the transition under the beam of the potential of an element from $V_1(x,y)$ to the positive equilibrium potential V_2. It is of fundamental importance, for, in this step, the picture signal is generated.

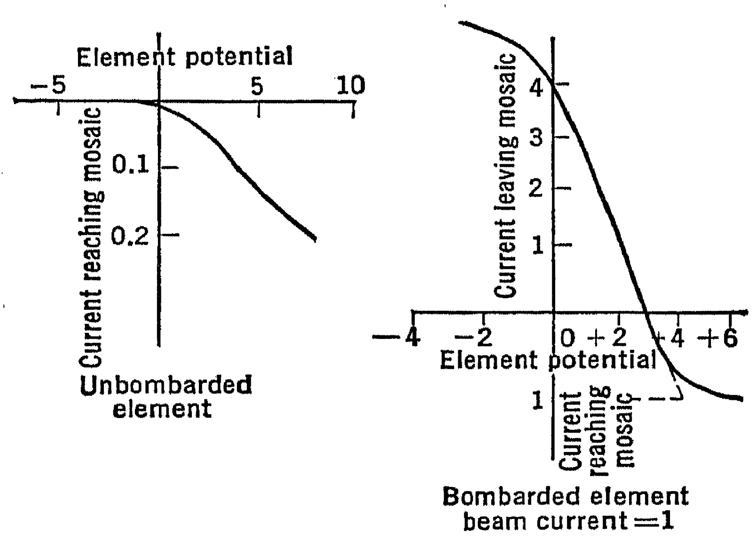


FIG. 10.16.—Current-Voltage Characteristics for an Element of the Mosaic.

The beam in sweeping over the mosaic acts like a resistive commutator in series with a source of positive potentials of 2 or 3 volts. The resistance in this case is determined by the current-voltage relation of the secondary electrons from the bombarded area, and will, of course, be non-ohmic. By means of small, caesiated silver electrodes inserted into the mosaic, it is possible to learn a little about the nature of this relation. Fig. 10.16 shows typical curves of the current to or from such an element as a function of voltage. It is found that the curve may be approximated over a large part of its range by an ohmic impedance whose value is $z = z_0/i_b$ where i_b is the beam current, and z_0 the coefficient of beam impedance having a value between 1 and 2 ohm-amperes. Although this relation is useful

for certain purposes, where an averaging over a picture cycle is involved, it is not a sufficiently good approximation for the determination of the instantaneous condition of a bombarded element. It is probable that the emission from an element is actually very nearly saturated (that is, most of the electrons leaving it return to other points of the mosaic rather than to the element under bombardment) until it is very close to the equilibrium potential, V_2.

For purposes of determining the change in potential ΔV of an element of area which is small compared with the size of the scanning spot, yet includes a large number of silver particles, during bombardment, let the saturated secondary emission current density be ρ_s and the capacity

$$C = C_0 ds.$$

If it is assumed that the secondary emission is saturated for the duration

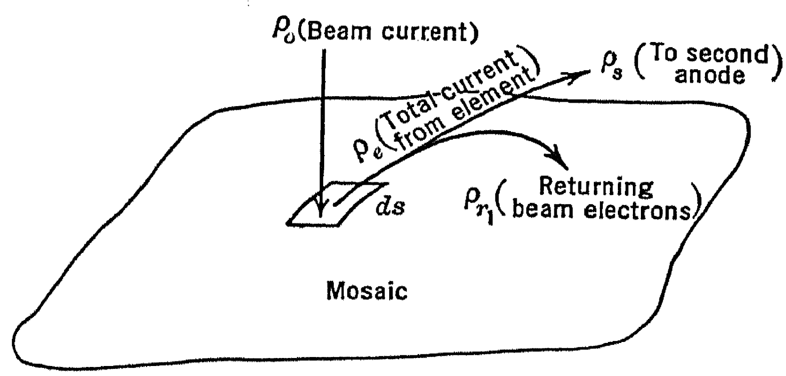


FIG. 10.17.—Current Components for an Element under the Beam.

of the time, T_e, the beam is on the element, the change in potential will be:

$$\Delta V = \frac{(\rho_s - \rho_b)ds}{C_0 ds} T_e. \tag{10.1}$$

For a beam current of 0.20 microampere, ρ_b has a value 5×10^{-4} amp/cm², and assuming a secondary-emission ratio of 4, $\rho_s - \rho_b = 1.5 \times 10^{-3}$. The element time, T_e, can be found by dividing the linear scanning velocity, v, into the width of the spot, h, which gives 1.25×10^{-7} second for the time the element is being bombarded. Introducing these values into Eq. 10.1, the maximum value for the change in voltage is around 2 volts.

Experimentally, it is found that, under the beam conditions described, the change in voltage $\Delta V = V_2 - V_1$ lies between 1 and 2 volts, which is consistent with the above.

The net current leaving the elementary area (see Fig. 10.17), under the conditions just described, will be:

$$di_e = -\left(C_0 \frac{dV}{dt}\right) ds.$$

Since the beam current reaching the area is $\rho_b ds$, the total current from the element is:

$$di_t = -\left(\rho_b + C_0\frac{dV}{dt}\right)ds.$$

This entire current will not reach the second anode because, on the average, as much current must reach the mosaic as leaves. Let $f(x,y,V)$ be the fraction of this current reaching the second anode; then the current to the second anode can be written as:

$$di_s = f(x,y,V)\left(\rho_b + C_0\frac{dV}{dt}\right)ds. \tag{10.2}$$

The instantaneous current reaching the second anode is given by the integral of the above expression over the area of the scanning spot. Since the current i_s carries the picture signal, the solution of Eq. 10.2 would be highly desirable. This integration, however, cannot be performed since the function $f(x,y,V)$ is unknown.

It is possible to make certain approximations which will permit the calculation of an average value for the redistribution function. The mosaic is an insulator and returns a current to the second anode equal to the beam current. Hence the average value of this function must be given by:

$$\bar{f} = \frac{i_b}{i_t} = \frac{i_b}{i_b + i_e}. \tag{10.3}$$

The average value of the element current, i_e, must be equal to the charge released by the beam divided by the time required for its release. On this basis, i_e is given by:

$$i_e = \int_{\text{spot area}} \frac{(V_2 - V_1)C_0 ds}{T_e}$$
$$= \frac{(V_2 - V_1)C_0 h^2}{h/v} = (V_2 - V_1)C_0 hv.$$

Substituting this into Eq. 10.3, the average value for the redistribution function will be:

$$\bar{f} = \frac{1}{1 + \dfrac{hvC_0(V_2 - V_1)}{i_b}}. \tag{10.4}$$

Experimental quantities representing the ordinary operating condi-

tions being taken, the value of the function is found to be about 0.25. This may be interpreted to mean that approximately a quarter of the total charge which leaves the mosaic reaches the second anode, the rest being returned to the mosaic in the form of a redistribution current. Since the variation in V_1 over the mosaic, under conditions of low light levels, is small, it is legitimate to treat f as independent of the charge image. Therefore, only about 25 per cent of the stored charge is available for producing the picture signal.

The returned signal, which constitutes about three-fourths of the net stored charge, can very easily be shown experimentally. The Iconoscope required has a mosaic whose sensitized surface is made in the ordinary way, but whose signal plate is divided into two halves, each of which is brought out on a separate lead. The video amplifier is connected to one half of the signal plate while the other half is grounded. If a light image is projected onto the grounded side of the mosaic, a negative of this image will be seen on the screen of the viewing tube. This image is due, of course, to the redistribution electrons returning to the mosaic carrying some of the signal back in the form of a negative picture. With normal connections, this negative signal tends to oppose the positive picture, so that the net output of the tube is only about 25 per cent of what might be expected if this redistribution phenomenon did not exist.

10.9. Line Sensitivity. Going back for a moment to the question of the photoemission of the mosaic, an examination of the potential distribution over the mosaic reveals that there is an abrupt potential rise at the line being scanned, and that the line just in front of it is subjected to a strong positive field which tends to saturate the photoemission. This means that the line in question is much more sensitive than any other part of the mosaic, and furthermore, even if "saturation" light levels are used so that an equilibrium potential image has been established, further photoemission can take place at this line. This phenomenon is known as line sensitivity and is the reason for the statement that the picture signal keeps on increasing even when the light level is increased above that required to form an equilibrium potential image.

The phenomenon of line sensitivity can be demonstrated very strikingly by the following simple experiment. The image from a continuously run moving-picture film (i.e., that obtained by removing the intermittent and shutter from a moving-picture projector) is projected onto the mosaic of an Iconoscope. The film is run at such a rate that the frame speed is equal to the field frequency and in a direction such that the image moves opposite to the vertical direction of scanning. Under these conditions the Iconoscope transmits a clear image of two frames of the

moving-picture film, although to the eye there appears to be only a blur of light on the mosaic.

Thus, it is evident that there are two sources of signal, one the stored charge on the entire mosaic surface, the second the sensitive line ahead of the scanning beam. At low light levels the surface sensitivity contributes the major portion of the signal, but at high illumination intensities, line sensitivity can contribute a large fraction.

10.10. Black Spot. Up to this point it has been assumed that the two terminal potentials, V_1 and V_2, had the same value at every point on an unilluminated mosaic. Actually, because both potentials are sensitive to the field conditions in their neighborhood, they are functions of the coordinates of the mosaic. Furthermore, the redistribution factor, f, is also dependent upon the location of the point under bombardment. As a consequence, there is a variation in the current reaching the second anode as the mosaic is scanned. This gives rise to a spurious signal which, if not compensated, produces an irregular shading over the picture. It is evident from the factors upon which this shading depends that it may be considered as divided into two parts. The first is a stored signal which depends upon $V_1(x,y)$; the second is an instantaneous effect depending upon the variation of $V_2(x,y)$ and $f(x,y)$ over the mosaic.

The instantaneous component of the spurious signal can be demonstrated experimentally if a metal plate is substituted in place of the usual mosaic. When such a plate is scanned in the ordinary way, it being maintained at a potential close to that of the second anode, a spurious signal is generated which is very similar to that produced by the Iconoscope. The mechanism of the generation of this signal, though not identical with that involved in the case of the mosaic, is quite similar. If the potential of the plate is made strongly positive or negative with respect to the second anode, so that the secondary-emission current reaching the second anode is either saturated or suppressed, the signal will disappear.

In the normal Iconoscope the magnitude of the spurious signal depends upon: (1) the non-uniformity of scanning, (2) the shape of the glass envelope and extent of exposed glass, (3) the beam current, and (4) the activation of the mosaic. It can be reduced by attention to items 1, 2, and 4, but it cannot be completely eliminated. With respect to its dependence upon beam current, it is found to increase more rapidly than the picture signal as the beam current is increased. For this reason, it is common practice to operate the Iconoscope with small beam currents (i.e., from 0.1 to 0.2 microampere) at the expense of some efficiency in signal generation.

In television transmission the spurious signal from the Iconoscope is compensated by means of an electrical connecting network, which intro-

duces a signal similar in shape to the shading signal but opposite in polarity into the video amplifier.

No mention has been made of space charge between the mosaic and second anode. Between these elements there necessarily exists an electron cloud, but under ordinary operating conditions its density does not become sufficient to create a potential minimum or a virtual cathode. Although the space cloud of electrons is incidental only, as far as the operation of the Iconoscope is concerned, nevertheless it has an effect on the paths of the redistribution electrons, and can introduce transit time effects observable under certain unusual operating conditions. Very close to the element under bombardment, when it is near its equilibrium potential, a virtual cathode is undoubtedly formed.

10.11. Performance of the Iconoscope. From the theory of operation just presented, the signal output from a standard Iconoscope falls below that expected from an ideal storage tube owing to lack of saturation of the photoemission and to redistribution losses. Assuming 20 per cent for the former and 25 per cent for the latter, the overall efficiency is expected to be in the neighborhood of 5 per cent. In spite of this inefficiency, the very great advantage resulting from the application of the storage principle makes the Iconoscope a very effective pickup tube.

On the basis of the mechanism predicated and the estimated efficiency, the output expected can be calculated. Let it be assumed that a small area of the mosaic is illuminated with an intensity of L lumens per square centimeter. Also the following additional data taken from the characteristics of the standard Iconoscope are necessary:

Area of mosaic.	$A = 100\ \text{cm}^2$
Photosensitivity.	$p = 15\ \mu\text{A/lumen}$
Overall efficiency.	$k = 0.05$
Coupling resistor.	$R = 10{,}000$ ohms

The increase in current reaching the second anode as the beam reaches the illuminated area will be given by the additional charge accumulated per unit area by the mosaic multiplied by the rate at which the beam traverses this area and the redistribution efficiency. In other words:

$$\begin{aligned} i_s &= (kpLT)h.v \\ &= k \cdot p \cdot L \cdot T \cdot \frac{H}{n} \cdot W \cdot \frac{n}{T} \\ &= kpLA. \end{aligned} \tag{10.5}$$

In this derivation, H and W are the height and width of the mosaic, n the number of lines, and T the picture period. The remaining symbols have already been defined.

The voltage appearing across the coupling resistor will be:

$$V_s = Ri_s = kpLAR.$$

Substituting the values given above, the signal voltage is found to be:

$$V_s = 0.75L,$$

where V_s is in volts and L in lumens per square centimeter, or, as is generally found more convenient, V_s is in millivolts and L in millilumens per square centimeter.

Actual response curves taken of good Iconoscopes have been found to have an initial slope of more than 1 millivolt per millilumen per cm². However, the agreement between the measured signal output curves and the response predicted is within the accuracy of the assumptions that were made.

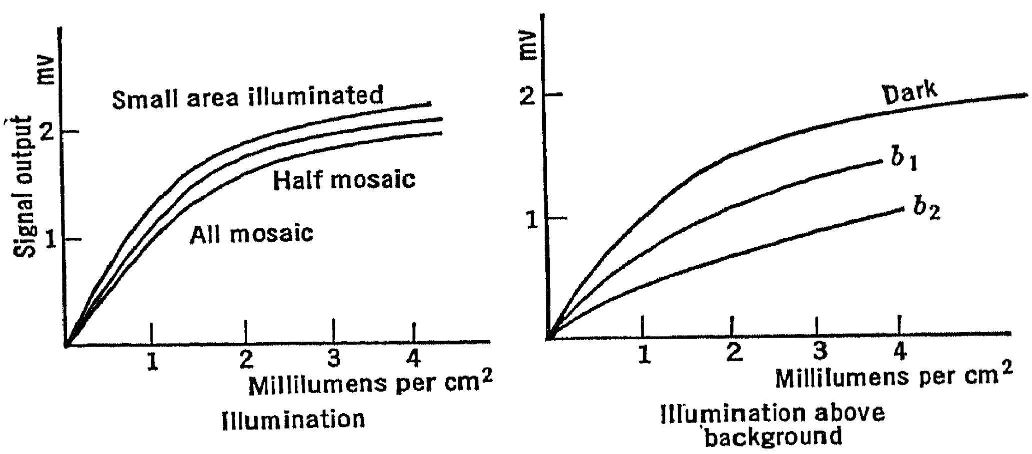


FIG. 10.18.—Response of Iconoscope under Various Conditions of Illumination.

The two sets of curves reproduced in Fig. 10.18 are typical of the response of the Iconoscope. Those shown on the left represent the signal output as a function of the intensity of illumination for three different sizes of area illuminated. The family of curves on the right was obtained by measuring the signal voltage generated across a 10,000-ohm coupling resistance when a spot of light occupying 1 per cent of the area of the mosaic was projected on the latter. In addition to the spot of light, a uniform background as indicated covered the entire mosaic. The highest curve a is the result of the measurement when the background illumination was zero. For this curve the initial slope corresponds to a sensitivity of 0.95 millivolt per millilumen. As the light intensity is increased, the slope of the curve becomes less. The shape of the response curve over a large part of its range is logarithmic, but departs from this at the two extremes of illumination. The curves b_1 and b_2 illustrate the effect of a background illumination. It is evident that, as the background illumina-

tion increases, the initial slope of the response curve decreases. Empirically, the initial slope D is related to the background illumination L_B as follows:

$$D = \frac{D_0}{1 + mL_B},$$

where D_0 is the slope for zero background and m is an empirical constant. It is interesting to note that this relation is derivable from the logarithmic form of the response curve. However, the magnitudes of the constants are functions of the areas illuminated by the spot and the background.

The color response of the Iconoscope depends upon the activation of the mosaic. By properly selecting the activation schedule, the response

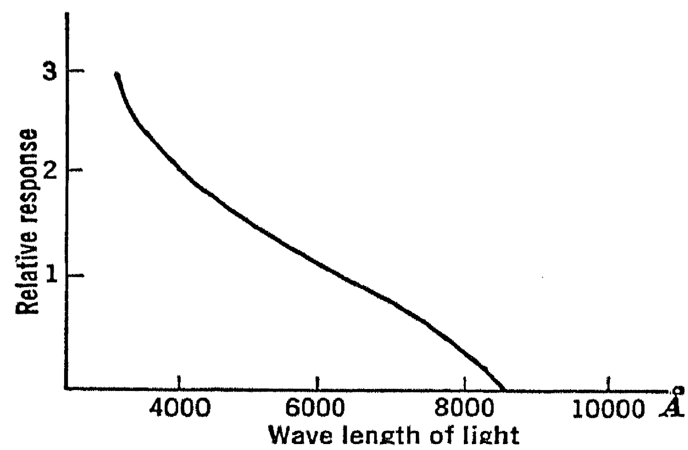


Fig. 10.19.—Spectral Response of an Iconoscope.

can be shifted to meet the operating requirements over a wide range. Fig. 10.19 shows the color response of a typical normal Iconoscope.

Since the mosaic of the Iconoscope is surfaced with silver particles which are very small compared with the beam size, the resolution which can be obtained is limited by the latter. The resolution of the tube is usually measured, as has been explained, by projecting onto the mosaic a test pattern consisting of groups of converging lines (see Fig. 10.9). With such a pattern it can be shown that the horizontal resolution of the normal Iconoscope is in the neighborhood of 700 or 800 lines. Since the video amplifier response extends to about 4 megacycles, a picture of this definition cannot be transmitted over the system. However, the high limiting resolution insures a small aperture defect over the working region.

10.12. Limiting Sensitivity. The question of the statistical fluctuation and its relation to the minimum light that can be used to produce a picture is of fundamental importance. Broadly speaking, these fluctua-

tions arise from the granular nature of electricity and are present in every electronic device and every circuit element having a resistive component. As has been shown, the Iconoscope generates a certain signal voltage across the coupling impedance. At the same time, there appears across this impedance a random fluctuating voltage due to the thermal agitation of the electrons in the coupling resistance. The video amplifier amplifies not only the picture signal but also the voltage fluctuations of the coupling resistance, or, to adopt the terminology of amplifier practice, the "noise." If the voltage generated by the picture signal becomes of the same order as the noise, the picture becomes lost in noise and is no longer intelligible. This sets a definite lower limit to the light which can be used to produce a picture from a given tube.

Before making any calculations on the sensitivity of the Iconoscope, it is necessary to assign quantitative values to the psychological effect of the picture-to-noise ratio. Tests have been made to determine the effect on the observer of various ratios of peak picture signal in an average picture to the root mean square noise. It has been found that, if the r.m.s. noise is equal to 30 per cent of the picture signal, the picture is still recognizable, but the definition is decreased and the picture tiring to watch. In addition, it was observed that, if the ratio remains constant but the picture amplitude is decreased, the noise becomes less objectionable; however, the effective resolution is not increased.

If the ratio of signal-to-noise is 10 to 1, a good picture can be obtained, but the noise is still very noticeable. Such a picture is usable and has fair entertainment value.

When the noise is reduced to less than 3 per cent of the picture signal it becomes practically unnoticeable and the picture may be considered excellent.*

Most of the calculations of sensitivity which follow will be based on an allowable signal-to-noise ratio of 10. Although this amount of noise would be greater than should be tolerated if conditions made it possible to avoid it, such a picture would, nevertheless, have reasonably good entertainment value and could be broadcast where program continuity made its transmission necessary.

The noise which must be considered in sensitivity calculations arises primarily from two sources: first, the coupling resistance, or impedance, between the Iconoscope and the video amplifier; and second, the first tube of the amplifier.

The thermal noise generated by a resistance is found to have the following characteristics. It appears as a voltage generated across the ends of the resistance, and is entirely independent of any currents flowing

* These tests were made with noise evenly distributed over the frequency band. If high-frequency noise predominates the percentage r.m.s. noise permissible is greater.

through it. The fluctuations are completely random, so that their energy is distributed uniformly over the frequency spectrum. Furthermore, the mean square voltage is proportional to the temperature of the resistance and to its magnitude or, to be more general, the magnitude of the resistive component of any impedance. All this can be formulated in the following equation:

$$\overline{e_n^2} = 4kTRF,$$

where k is the Boltzmann constant and T the absolute temperature, the product $4kT$ having a value of 1.6×10^{-20} joule at room temperature. R is the resistive component of the impedance, F the frequency band under consideration, and $\overline{e_n^2}$ the mean square noise voltage.

In addition to the noise generated by the coupling impedance, noise is generated by the amplifier itself. This noise is primarily due to the first

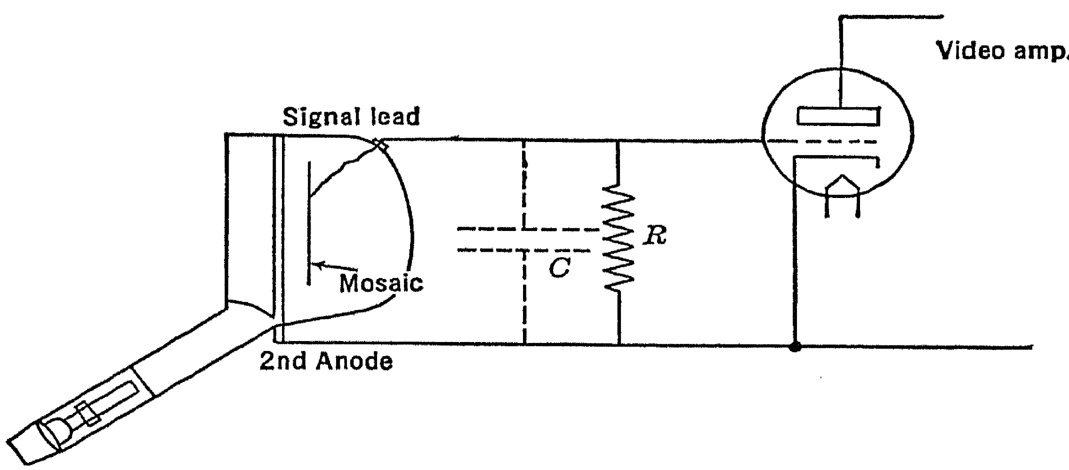


FIG. 10.20.—Simple Iconoscope Coupling Circuit.

tube of the amplifier, but the coupling impedance between the first and second stage may also contribute an appreciable amount. The noise produced by the amplifier depends, of course, upon the tubes used. A noise characteristic corresponding to a 500- or 1000-ohm resistor in the grid circuit of the first tube will represent a typical, well-built amplifier.

Heretofore a 10,000-ohm coupling resistance between the Iconoscope and the picture amplifier has been assumed in discussing the performance of the tube. In actual television practice the coupling cannot be as simple as this, because the capacity of the Iconoscope and amplifier input circuit causes an attenuation of the high-frequency components of the picture signal. However, in order to estimate the sensitivity of the pickup tube, the simple coupling circuit consisting of a 10,000-ohm resistance will be retained and the attenuation of the upper frequencies neglected. Details of the more complicated circuits needed to maintain a uniform frequency response will be given in Chapter 14. There it will be seen that the limiting sensitivity that can be obtained with the corrected circuit

networks actually used does not differ greatly from that given under the present simplification.

The mean square noise voltage generated by the 10,000-ohm resistor over a 4-megacycle frequency band is:

$$\begin{aligned}\overline{V_N{}^2} &= 1.6 \times 10^{-20} \times 10^4 \times 4 \times 10^6 \\ &= 6.4 \times 10^{-10}.\end{aligned}$$

The Iconoscope, being a high-impedance current generator (actually its zero-frequency impedance is about 5 megohms when a 0.2-microampere beam is used) produces a signal voltage across the coupling resistor which is given by:

$$V_s = pkALR \text{ volts},$$

or, using the values from page 291, by:

$$V_s = 0.75\ L \text{ volts}.$$

The ratio S of signal to root mean square noise is therefore

$$S = \frac{V_s}{\sqrt{\overline{V_N{}^2}}} = 3 \times 10^4 L.$$

Assuming that the signal-to-noise ratio cannot be less than 10, the lower limit of light which will give a usable picture signal is

$$L = 3 \times 10^{-4} \text{ lumen/cm}^2$$

Similarly, to obtain a picture with a signal-to-noise ratio such that the noise is less than 3 per cent of the picture signal, the illumination on the mosaic must be at least a millilumen per square centimeter.

The order of magnitude of the measured sensitivity agrees fairly well with these calculations. Tests on a good Iconoscope have been made indicating that a satisfactory image can be obtained of scenes having a brightness of 15 candles per square foot, using a lens with an $f/4.5$ aperture. The illumination on the mosaic, taking account of reflection losses in the lens, will then be about 0.4 millilumen per square centimeter.

Usually the illumination for actual television transmission is made, if possible, at least three times this amount, to insure that the picture signal will be well above all spurious signals. This means that a surface brightness of 50 candles per square foot is desirable. A reflection coefficient of 25 per cent being assumed, the incident illumination of the scene should be $\dfrac{50 \times \pi}{0.25} \cong 600$ foot-candles. Of course, a transmittable picture can be had with less than 20 per cent of this illumination, but the figure given above can be taken as the optimum.

TABLE 10.1

Scene	Location	Time (E.S.T.)	Date	Weather	Surface Brilliance (candles/ft²)
Sixth Avenue	New York, N. Y.	9:30 A.M.	4–25–35	Clear	6½
Sixth Avenue	New York, N. Y.	1:15 P.M.	4–25–35	Overcast	40
Times Square	New York, N. Y.	1:30 P.M.	11– 6–34	Light rain	40
Parade	East Orange, N. J.	10:30 A.M.	11–29–34	Light rain	40 to 60
Street	Rockland, Me.	1:15 P.M.	7– 5–36	Overcast	100
Street	Warrenton, N. C.	3:15 P.M.	6–30–35	Clear	130
Street	Harrison, N. J.	3:30 P.M.	8–15–34	Hazy	130
Street	Harrison, N. J.	9:30 A.M.	8–15–34	Rain	16
River	New York, N. Y.	2:30 P.M.	10–24–35	Hazy	50
River	Pennsville, Del.	1:30 P.M.	6–29–35	Hazy	350
Bay	Cape Charles, Va.	10:00 A.M.	6–30–35	Clear	250
Beach	Atlantic City, N. J.	2:00 P.M.	8–18–34	Hazy	500
Football game	New York, N. Y.	1st quarter	11–17–34	Clear	55
		2nd quarter		Hazy	50
		3rd quarter		Hazy	27
		4th quarter			16
		End			2
Baseball game	New York, N. Y.	1:00 to 3:40	9– 8–35	Clear	70 to 100
Snow bank	Harrison, N. J.	10:00 A.M.	1–24–35	Bright sunshine	700
Open field	Bethel, N. C.	3:45 P.M.	7– 1–35	Severe thunderstorm	2

In order to give an idea of the brightness encountered in typical outdoor scenes, Table 10.1 is reproduced from a paper of Iams, Janes, and Hickok.* Since it can be assumed that the Iconoscope has sufficient sensitivity to transmit a usable picture at least down to brightness of 10 candles per square foot, the versatility of the tube will be evident.

If the walls of a normal Iconoscope are photosensitive and illuminated with a small amount of light, the photoemission from the bulb so alters the conditions in the vicinity of the mosaic that the working sensitivity is about doubled. This expedient, which is of considerable practical importance, was first introduced by the Electrical and Musical Industries of England, and has since been widely adopted.†

10.13. Depth of Focus. The depth of focus of the objective used with an Iconoscope is an important consideration. In fact, if it were not for this limitation, the Iconoscope with very large-aperture objectives would be sufficiently sensitive to transmit a picture under all practical conditions, both indoors and out. However, the aperture can be increased only at the expense of depth of focus. In the photographic camera, depth of focus is obtained by means of high-aperture, short-focal-length lenses. This procedure permits decreasing the diameter of the lens without decreasing the intensity of illumination on the photographic plate. Since

* See Iams, Janes, and Hickok, reference 7.

† See Janes and Hickok, reference 10.

depth of focus in the object field is dependent only upon the diameter of the lens and not upon the numerical aperture, it can be increased without loss in sensitivity.

It has been shown above that in the Iconoscope the response is proportional not only to the intensity of illumination but also to the mosaic area. Stated in another way, the response is proportional to the total light falling on the mosaic. This light is in turn proportional to the absolute diameter of the lens rather than to its numerical aperture. Therefore, it is not possible to follow photographic practice and increase the depth of focus by using a miniature Iconoscope and a short-focal-length lens. The conclusion is that, for a given sensitivity of the Iconoscope and a given illumination of the scene transmitted, the signal obtained is inversely proportional to the square of the depth of focus in the object field.

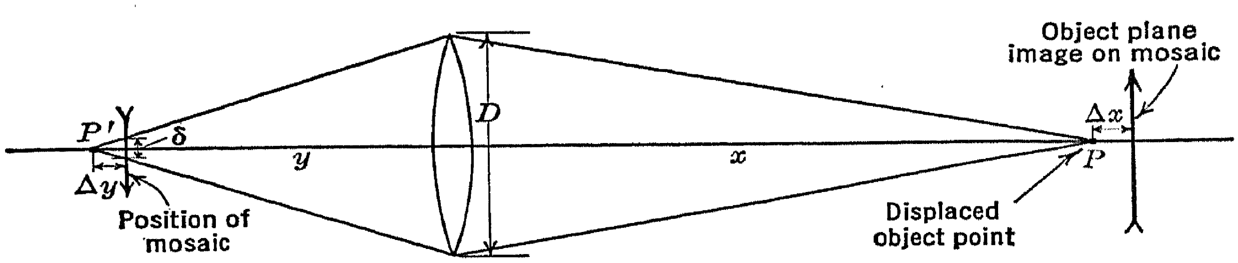


Fig. 10.21.—Determination of Depth of Focus.

The relation between depth of focus and the lens diameter can be shown as follows. Referring to Fig. 10.21, the object and image distances are x and y, respectively, and the focal length of the lens is f. These quantities are related by the equation:

$$y = \frac{fx}{x - f}.$$

By differentiation, a change Δx in the object distance, x, is found to produce a change in the image distance:

$$\Delta y = -\frac{y^2}{x^2}\Delta x.$$

This will produce a circle of diffusion of diameter δ in an image at y, with:

$$\delta = \frac{D}{y}\Delta y = D\frac{y}{x}\frac{\Delta x}{x},$$

where D is the diameter of the lens. For a given object, the effective resolution will be equal to the diameter of the circle of diffusion divided

by the magnification. Since the magnification, m, is given by y/x, the quantity to be determined, that is δ/m, can be expressed as follows:

$$\frac{\delta}{m} = D\frac{\Delta x}{x}.$$

Therefore, the depth of focus in the object field $\Delta x/x$ is dependent only upon the diameter of the lens, D.

10.14. Pickups for Motion-Picture Film. So far, no mention has been made of the transmission from moving-picture films. Two methods of reproduction will be discussed. The first makes use of a continuous motion projector. Such a projector is arranged so that the light on the screen remains constant and the consecutive frames fade smoothly into one another. The operation of the Iconoscope with such a projector is exactly the same as in studio or outdoor pickup. Furthermore, the frame frequency of the film need bear no relation to the scanning frequency of the tube. A second method, somewhat more complicated but nevertheless found in practice to be quite satisfactory, is to project the picture onto the mosaic during the return time of the beam. An ordinary moving-picture projector, using a standard film which is run at 24 frames per second, is equipped with a shutter, usually in the form of a rotating, apertured disk, which allows the light image to pass 60 times a second for a period of about 1/800 second. With this arrangement it will be obvious that images from consecutive frames will pass through the shutter alternately two and three times. The scanning beam is synchronized with the shutter in such a way that the image is flashed on the mosaic only during the time between the end of one scanning pattern and the start of the next, while the beam current is biased to cutoff. The light image flashed on the mosaic causes a potential image to be formed in the manner described in the section discussing the theory of the Iconoscope. This potential image gives rise to the picture signal when the unilluminated mosaic is then scanned. A reproduced picture thus formed differs in no way from the ordinary picture in appearance. Because of the short duration of the light image on the mosaic, the instantaneous illumination must be high, but this presents no particular difficulty since the light produced by an ordinary projector is ample.

Further discussion of the practical aspects of moving pictures is to be found in Chapter 18, describing the RCA television project.

10.15. Summary. To summarize the performance of the Iconoscope as a pickup tube, it may be said that, although it is not without its limitations, it is nevertheless capable of reproducing most scenes encountered in practice. Where an illumination of 1000 to 2000 foot-candles is available, the tube imposes essentially no limitation on the performance of

the system. Such illuminations are always available in the studio, and are to be found outdoors throughout the year on clear days between the hours of 9:00 A.M. and 3:00 P.M., at latitudes of 42°N. In summer this illumination reaches a very high value at noon, that is, around 9000 foot-candles, and illuminations of 1000 or 2000 foot-candles are available from 6:00 A.M. to 6:00 P.M. Of course, clouds or shade reduce the available illumination, so that ideal transmission is not always possible during the hours indicated. At low levels of light the pickup tube limits the performance by making impractical depths of focus necessary, or introducing excessive spurious signal or noise. However, over the range of illumination from 1000 to 50 foot-candles a satisfactory picture can be reproduced. Below this range the picture rapidly loses entertainment value.

REFERENCES

1. FRITZ SCHRÖTER (editor), "Fernsehen," Julius Springer, Berlin, 1937.
2. BUSCH and BRÜCHE (editors), "Beiträge zur Elektronenoptik," J. A. Barth, Leipzig, 1937.
3. V. K. ZWORYKIN, "The Iconoscope—A Modern Version of the Electric Eye," *Proc. I. R. E.*, Vol. 22, pp. 16–32, January, 1934.
4. V. I. KRASOVSKY, "On Light Storing Devices," I. E. S. T. (*Izvestia Elektro-Promishlenosti Slabovo Toka*), No. 2, pp. 24–34, February, 1936.
5. R. URTEL, "The Operation of Electron Beam Scanning Devices with Storage," *Z. Hochfreg. Elektroakust.*, Vol. 48, pp. 150–155, November, 1936.
6. V. K. ZWORYKIN, G. A. MORTON, and L. E. FLORY, "Theory and Performance of the Iconoscope," *Proc. I. R. E.*, Vol. 25, pp. 1071–1092, August, 1937.
7. H. IAMS, R. B. JANES, and W. H. HICKOK, "The Brightness of Outdoor Scenes and its Relation to Television Transmission," *Proc. I. R. E.*, Vol. 25, pp. 1034–1047, August, 1937.
8. W. HEIMANN AND K. WEMHEUER, "Operation of the Electron Beam Picture Scanner," *Elek. Nach.-Tech.*, Vol. 15, pp. 1–9, June, 1938.
9. M. KNOLL, "The Significance of the 'Redistribution Electron Effect' for the Operation of Picture Scanning Tubes," *Z. tech. Physik*, Vol. 19, pp. 307–313, October, 1938.
10. R. B. JANES and W. H. HICKOK, "Recent Improvements in the Design and Characteristics of the Iconoscope," *Proc. I. R. E.*, Vol. 27, pp. 535–540, September, 1939.

CHAPTER 11

THE ICONOSCOPE, Continued

In the preceding chapter the theory and performance of the normal Iconoscope were discussed in some detail. Possible methods of improving the Iconoscope, particularly in respect to sensitivity, will be the subject matter of this chapter.

Even though the Iconoscope has, under quite exhaustive tests both in the laboratory and in the field, proved itself to be a very satisfactory pickup device, nevertheless a real improvement in picture will be the result of an increase in sensitivity. Increased sensitivity will permit the use of lenses having greater depth of focus, will extend the possibilities of obtaining an optimum picture in outdoor transmission, and will improve studio working conditions.

The problem of increasing the sensitivity of the Iconoscope may be approached in at least three different ways; first, by increasing the overall efficiency of the mosaic; second, by keeping the signal output the same and reducing the noise generated; and finally, by increasing the quantity of charge acquired by the mosaic per unit light flux. This list does not, of course, exhaust all possibilities, but it will at least indicate the extent of the field open for research.

The remaining sections relating to the Iconoscope will take up these approaches in turn and will describe some of the possible solutions of the problem of increasing the sensitivity of the pickup tube. It should be remembered that most of the developments described in what follows represent work that is being carried on in various research laboratories, such as that at RCA Radiotron, RCA Victor, and many others, and should be considered as presenting a résumé of the experimental field as of 1939 rather than a description of finished products. Furthermore, the theories and conclusions offered below are purely tentative and may become obsolete or be found invalid as knowledge in the field of television is advanced.

METHODS OF INCREASING MOSAIC EFFICIENCY

The most immediately apparent way of increasing the effectiveness of the mosaic is by improving the photosensitivity of its surface. A great deal of work has been done along these lines, with the result that in the

last two or three years the photosensitivity has been increased from around 5 to 7 microamperes per lumen to 12 or 15 microamperes per lumen, and the signal output from 0.4 to 1.0 millivolt per millilumen per square centimeter. Continued research on this problem will unquestionably yield further improvements both in color response and sensitivity.

The effectiveness of the mosaic can also be increased by eliminating, or reducing, the losses mentioned in the discussion of the normal Iconoscope. The most serious of these losses are the consequence of (1) unsaturated photoemission, and (2) redistribution of electrons.

11.1. The Two-Sided Mosaic. The generation of the picture signal from the scanned mosaic of a normal Iconoscope is possible only when the elements can be driven to an equilibrium potential such that the effective secondary-emission ratio is unity. Therefore, to serve as a source

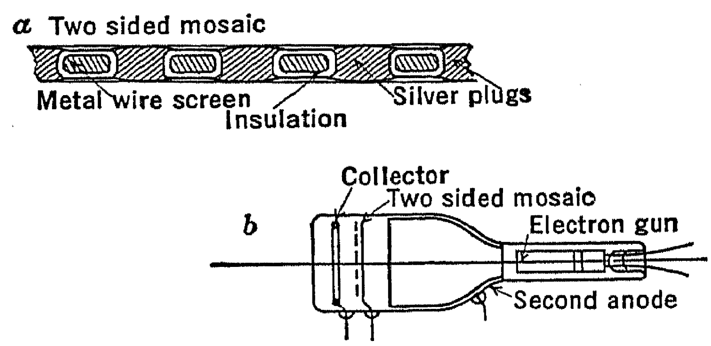


FIG. 11.1.—Iconoscope with Two-Sided Mosaic.

of picture signal the potential of the elements cannot differ greatly from that of the electrode which collects the secondary electrons emitted by them. One method of effecting saturated photoemission while still retaining unsaturated secondary emission is to provide a collecting electrode for each function and to separate the two functions. This can be done by constructing the mosaic in such a way that the photoemission takes place from one side and the beam scans the other.*

A two-sided mosaic to be so operated may be made by coating a fine-mesh metal screen with an insulating layer and filling the interstices with metallic plugs which are photosensitized on the side away from the beam. Fig. 11.1*a* shows such a mosaic in cross-section.

Screens suitable for the purpose can be woven of fine wire, or they can be made by electrodeposition, similarly to half-tone screens. The size of mesh depends upon the resolution desired and upon the overall size of the mosaic. It should be such that at least four elements are under the beam at all times. Fewer elements than this will give the picture a grainy appearance which is highly undesirable. Experience shows that 150 to

* See Zworykin, Morton, and Flory, reference 1.

200 wires per inch permits satisfactory resolution without requiring an excessively large mosaic. For a maximum coverage by photosensitive elements, the ratio of open area to metal of the screen should be as large as possible.

The insulating material for coating the screen should have the following qualifications: it must be inert to the chemical action of the activating agent, have a low vapor pressure, be able to withstand the temperature encountered in bake required for the outgassing of the tube during exhaust, adhere firmly to the screen without cracking or flaking off during the heating, and have high specific resistance. Vitreous enamels, lacquers, or ceramic materials can all be used as insulation for the screen. In applying the insulating layer, great care should be taken to avoid pinholes, cracks, or other openings, especially around the inside surfaces of the holes, as these imperfections will cause strong spurious signals to be generated.

The metal plugs which serve as the picture elements can be made in a variety of ways. One is to fill the holes with a paste made of silver oxide suspended in a binder which can be driven off by heat. When the holes are thus filled, the mosaic is heated above the reduction temperature of the silver oxide, leaving silver plugs in the interstices.

The mosaic is mounted in the blank so that the tube axis coincides with a normal from the center of the mosaic. The arrangement of the tube is shown in Fig. 11.1*b*. The end of the tube away from the gun serves as the window through which the optical image is projected. Just inside this window is a conducting ring of sufficient diameter so that it does not interfere with projected pictures. This ring is the collector for the photoelectrons from the mosaic. The mosaic is mounted at a convenient distance back of the window. The side of the mosaic facing the window is subjected to an activation schedule similar to that used in the normal Iconoscope so that each element is coated with a photosensitive Cs-AgO-Ag film. For a signal plate the underlying metallic screen is used, it being connected to a signal lead which is brought out through the wall of the tube. The gun is of purely conventional design, mounted in a long thin arm, over which an ordinary deflecting yoke can be slipped. It is not necessary to place the gun at an angle to the screen since the light image is projected onto the opposite side, where it is free from possible interference from the gun. The second anode consists of a metallic film coated on the inside of the arm. This coating is extended into the tube almost to the mosaic, so that it also fulfills the function of collector of secondary electrons from the beam. The mosaic elements are maintained at a potential which is within a few volts of that of the second anode by the action of the scanning beam. Thus, if the photoelectric collector ring is

made 50 or more volts positive with respect to the second anode, ample field will exist on the photosensitized side of the mosaic to insure collection of all electrons released under the action of the light image.

This type of tube operates with a saturated photoelectric emission, but still suffers from the same loss of sensitivity due to the redistribution of beam secondary electrons.

The sensitivity at low light levels is several times that of a normal Iconoscope having a mosaic of equal size, capacity, and saturated photosensitivity, but owing to redistribution losses the saturated signal output is not much greater. Furthermore, the same spurious signal and shading consequent on redistribution will exist.

The chief obstacle in the way of the adoption of the two-sided mosaic is a purely technical one. The physical difficulty of building such a mosaic so that it is free from all blemishes such as those due to pinholes, elements short-circuited to the signal plate, variation in secondary emission from the surface of the insulating layer, and leakage through the insulation is enormous. Many problems in connection with its practical routine construction still remain to be solved.

11.2. The Barrier-Grid Mosaic. Where the principle used in the normal Iconoscope to attain element equilibrium under the beam is to be retained, secondary electrons returning to the mosaic are essential. If redistribution losses are to be avoided it is necessary to provide for the complete isolation of each individual element. A so-called barrier grid mosaic is one means for accomplishing this end. This mosaic is double-sided, similar to the one just described, but in addition a conducting coating is provided on the beam side, such that every element is encircled by, but not in contact with, this shield. The construction of this mosaic can be seen from Fig. 11.2*a*. There are obviously many ways of forming such a shield. To mention one, a metal film may be evaporated on the insulated mesh, suitable precautions having been taken to prevent the metal from getting into the holes of the screen.

The mosaic is mounted in the tube in the same way as the simple two-sided screen. The conducting shield faces the gun, and the opposite surface of the mosaic is photosensitized as before. The connections are the same as used for the first described except that the metallic shield, or barrier grid, is made 45 volts negative with respect to the second anode. A circuit diagram for the tube is shown in Fig. 11.2*b*.

When the tube is in operation the equilibrium potential of the elements under the beam is slightly positive with respect to the barrier grid. This potential is such that the electron current escaping from the element bombarded just equals the beam current. Because of the field existing between the barrier grid and the second anode, any electron which is

emitted with sufficient velocity to get an appreciable distance from the element is drawn to the second anode. Thus, all electrons which are emitted by the element bombarded must either return to the element itself or go to the second anode, the field of the shield eliminating all other possibilities. This eliminates completely the redistribution phenomenon and, with it, the loss of efficiency and spurious signal.

Tubes using this type of mosaic have more than 50 per cent of the ideal storage efficiency calculated from the saturated photoemission. Furthermore, as would be expected, the saturated signal output is extremely high. Instead of the usual 1 to 2 millivolts saturated signal generated by the normal Iconoscope, signals of several hundred millivolts are produced.

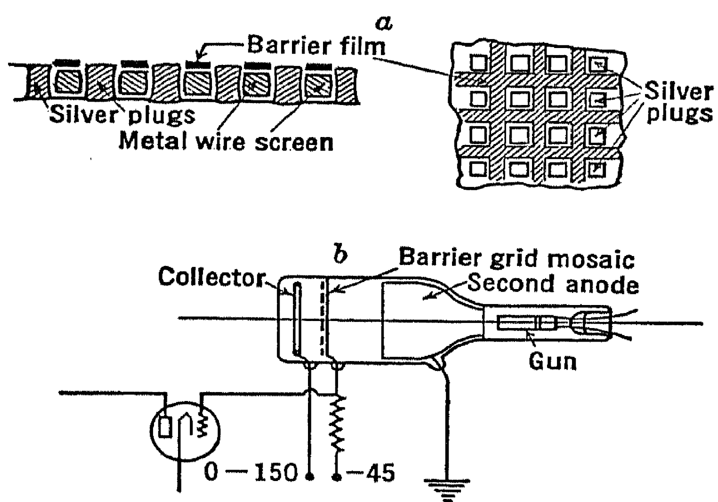


Fig. 11.2.—Barrier-Grid Iconoscope.

In discussing the simple, two-sided mosaic it was pointed out that the practical difficulties in the way of its construction were many. It is evident that even more serious ones apply to the barrier grid. Before this mosaic can be considered ready to take its place as part of a practical tube, a great deal of experimental work must be done toward its perfection.

11.3. The Conductive Mosaic. Besides the two double-sided mosaics just described, another possibility of improving the efficiency of the screen is through the use of a single-sided mosaic having a partially conducting dielectric layer instead of the usual high insulation.* Although a great deal of work has been done along this line, the results are too ambiguous to enable any positive conclusions to be formed. There are theoretical reasons which indicate that it is possible to increase the saturation of the photoemission while leaving the secondary emission of the

* See Iams and Rose, reference 2; Krawinkel, Kronjäger, and Salow, reference 11.

beam sufficiently unsaturated to perform its function of bringing the elements to an equilibrium. This leads to an increase in both saturated signal output and the sensitivity at low light levels. The extent of this increase is not readily calculable.

The use of a non-ohmic conductor instead of an insulator as the dielectric medium offers even greater possibilities. There are experimental indications that results comparable to those using a barrier grid mosaic are possible.

11.4. Low-Velocity Scanning. Another interesting line of approach, which has been the subject of investigation for the past several years, is the use of a scanning beam with a bombarding voltage sufficiently low so that the secondary-emission ratio is less than unity. The elements are driven to cathode potential for their equilibrium, at which potential the beam electrons can no longer reach them. Elements which are illuminated become slightly positive as a result of stored charge. When they are bombarded by the scanning beam this charge is released and the elements are driven again to cathode potential. Since an equilibrium is established between the mosaic and cathode, independent of the second anode potential, both photoemission and secondary emission can be saturated. This, of course, not only greatly increases the efficiency of the tube, but also eliminates the "black spot." The "Orthicon," a magnetic low-velocity Iconoscope with cycloidal deflection, is already in practical service.*

11.5. The High-Voltage Iconoscope. It was pointed out in Chapter 1 and again in Chapter 2 that the secondary-emission ratio of a bombarded surface decreases with increasing velocity of the impinging electrons after a certain optimum velocity has been exceeded. If the bombarding voltage is high enough the ratio falls below unity. The voltage for which the ratio is unity has been termed the break-point potential. If an insulated surface, such as a mosaic, is scanned with a high-voltage beam it will tend to assume the break-point potential, irrespective of the amount by which the potential between the cathode and the second anode exceeds the break-point potential. An elementary area of a mosaic operated under this condition, which as a result of incident light has been made positive with respect to the break-point potential, will be returned to the break-point potential when traversed by the beam. Thus a picture signal having the same sense as that from an ordinary Iconoscope is generated. Under this mode of operation, photoemission and secondary emission are saturated.

The beam impedance, as defined in Chapter 10, depends upon the beam current and the slope of the secondary-emission characteristic. Since the slope is small, an impedance low enough to discharge the mosaic in one picture period can be obtained only if the beam current is high.

* See Rose and Iams, reference 3.

11.6. The Image-Amplifier Iconoscope. A mosaic based upon the Iconoscope storage principle can be used as a grid to control a low-velocity scanning beam or the secondary emission from a high-velocity beam. The operation of this type of tube is as follows. A mosaic consisting of a perforated screen having insulated photoelectric rings or elements surrounding each hole on one side is mounted so that the opposite side is exposed to a low-velocity scanning beam. When an optical image is projected on the photosensitive side of the grid, the emitting areas are charged positive in proportion to the light intensity. The potential of the second anode on the scanned side of the mosaic, and that of a collector on the opposite side, are so adjusted that part of the electron beam returns to the second anode while the remainder goes through the screen to the collector. The fraction of the current reaching the collector for any position of the beam will be a function of the potential of elements surrounding the openings at that point. Thus, the current to the collecting element will be controlled by the light image on the mosaic grid and a picture signal will be generated. Provision, of course, must be made for restoring the control elements to an equilibrium potential.

As an alternative to a low-velocity beam, the mosaic grid may be scanned with a high-velocity beam which generates low-velocity secondary electrons upon striking the screen. The fraction of these low-velocity electrons passing through the screen is controlled by the stored photoelectric charge as before. To prevent the direct beam from penetrating the screen, the mosaic may be scanned from an angle, so that seen from the gun the screen wires appear to overlap one another. By properly adjusting the potentials in the tube the scattered electrons passing through the screen under the beam can be made to bring the elements to equilibrium.

The image amplifier Iconoscope has been investigated for a number of years at the RCA Manufacturing Company Laboratories and elsewhere, as it appears to have many interesting properties. Experimental models have been made which generate a fairly satisfactory picture, but many difficulties, some of them rather basic, must be overcome before it reaches the practical perfection of the normal Iconoscope. It is interesting to note that the limiting sensitivity of tubes of this class is about the same as that for the multiplier Iconoscope and the image Iconoscope, which are described later in this chapter.

THE SUPPRESSION OF NOISE

It has already been pointed out that the maximum sensitivity of the normal Iconoscope is determined by the noise generated in the amplifier and coupling impedance.

11.7. Signal-Multiplier Iconoscope; Theory. The secondary-emission

multiplier offers a possibility of amplifying the picture signal to a point where it is well above any amplifier noise before any coupling impedance is necessary.

To effect this form of amplification use is made of the fact that the effective circuit for the signal current from the Iconoscope is completed through secondary emission from the mosaic. In other words, if a coupling resistor and amplifier are connected to an electrode which collects the secondary electrons from the mosaic, a signal—albeit opposite in polarity—can be obtained which is equally as great as that from the signal plate. The electrons, instead of being collected from the mosaic directly, can be led into a secondary-emission multiplier from which the signal is obtained, thus eliminating the noise generated by the conventional amplifier circuit.*

However, this does not mean that unlimited sensitivity can be obtained. The noise produced by the shot effect in the scanning beam now determines the maximum sensitivity. Shot noise in the current from thermionic emitters such as are used in the gun was treated in Chapter 1. It was shown that the mean square current fluctuation in the emission from any temperature-limited cathode is given by the following relation:

$$\overline{i_N{}^2} = K_2 F i,$$

where K_2 is the shot coefficient having a value of 32×10^{-20} coulomb, F the frequency band over which observations are made, and i the current emitted by the cathode.

Since the beam impinges on a secondary emitting surface, the beam noise will at least be doubled. Space charge both in the gun and at the mosaic, on the other hand, will reduce the current fluctuations. However, for the purposes of the present calculation, it will be assumed that the mean square fluctuation in current leaving the mosaic will be double the shot noise in the beam. In other words, the mean square noise entering the multiplier will be:

$$\overline{i_N{}^2} = 2K_2 F i_b. \tag{11.1}$$

On this basis, the noise output from the multiplier,† if n stages are assumed with a gain of B per stage, will be given within the accuracy of the present calculation by:

$$\overline{I_N{}^2} = K_2 F \frac{2B^{2n+1} - B^{2n}}{B - 1} i_b. \tag{11.2}$$

* See Zworykin, Morton, and Flory, reference 1.

† See Zworykin, Morton, and Malter, reference 4.

The signal current entering the multiplier will be, by Eq. 10.5,

$$i_s = pkAL.$$

The squared signal output from the multiplier will then be:

$$I_s^2 = B^{2n}p^2k^2A^2L^2. \tag{11.3}$$

Dividing Eq. 11.3 by Eq. 11.2, the signal-to-noise ratio becomes:

$$S^2 = \frac{B^{2n}p^2k^2A^2L^2}{K_2F\left(\dfrac{2B^{2n+1} - B^{2n}}{B - 1}\right)i_b}$$
$$= \frac{p^2k^2A^2}{K_2F}\cdot\frac{B - 1}{2B - 1}\cdot\frac{L^2}{i_b}. \tag{11.4}$$

Evaluating the above expression for the Iconoscope data given in Chapter 10, page 291, and assuming a multiplier gain of 6 per stage and a frequency band of 4 megacycles:

$$S \cong 50\frac{L}{\sqrt{i_b}}. \tag{11.5}$$

If, from this formula, L is determined for a beam current of 0.2 microampere and a signal-to-noise ratio of 30, the minimum light for a good picture will be found to be about one-third of that required by the conventional system.

The point of interest to be noticed in Eq. 11.4 is that, if the beam current can be decreased without loss of efficiency of operation, the signal-to-noise ratio can be increased.

In a signal multiplier Iconoscope using an ordinary mosaic the beam current cannot be reduced without serious loss in efficiency. This is because, first, the average potential of the mosaic is more positive, and, second, the impedance of the beam becomes so high that the elements do not come to equilibrium in the time taken by the beam to cover them. The last condition means that the potential image will not be completely removed in a picture period, and moving images will appear blurred on the viewing screen.

To reduce the beam without introducing these effects, it is necessary to reduce the specific capacity of the mosaic at the same time. A convenient method of doing this is to increase the thickness of the dielectric between the sensitized surface and the signal plate.

It is of interest to estimate the performance which might be expected when the process is carried to a limit, assuming the same efficiency that exists in the normal Iconoscope. The lower limit of beam current is de-

termined by the photoemission on the mosaic, and is consequently a function of the incident light.

Assuming that the minimum beam is just equal to the effective photoemission, the beam current is:

$$i_b = k'pAL,$$

where k' is the fraction of the saturated photoemission actually available, allowing for the lack of saturation on the mosaic. Substituting this value in the previously derived Eq. 11.4:

$$S^2 = \frac{p^2k^2A^2}{K_2F}\frac{B-1}{2B-1}\frac{L^2}{LpAk'}$$

or:

$$S = \sqrt{\frac{pk^2A(B-1)}{k'K_2F(2B-1)}}\sqrt{L}. \tag{11.6}$$

Evaluating this for the case where $S = 10$ and $k/k' = 0.5$, the minimum illumination on the mosaic is:

$$L = 6 \times 10^{-6} \text{ lumen/cm}^2$$

The value is about 2 per cent of the theoretical lower limit of illumination for the conventional circuit and at most 1/20 of that required in practice.

11.8. Design and Construction of the Signal-Multiplier Iconoscope. In order to make full use of the secondary-emission multiplier it is necessary that it collect all those secondary electrons from the mosaic which would normally go to the second anode. This is one of the most difficult problems in the design of the signal-multiplier Iconoscope.

It is evident from the mechanism of operation of the Iconoscope that any electrode introduced into the Iconoscope which is made positive with respect to the second anode will cause the mosaic surface to rise in potential until it reaches approximately the same value as that of the electrode. If the potential of the electrode is not very different from that of the second anode, only part of the electrons from the mosaic may reach the electrode, the fraction reaching the electrode being different for different parts of the mosaic. In other words, the collection of signal electrons by this electrode tends to be non-uniform. If, on the other hand, the electrode is made very positive with respect to the second anode, all secondary electrons from the mosaic are collected. However, the fields resulting from the voltage difference between the electrode and the second anode will so upset the redistribution of electrons that very serious spurious signals are introduced. A satisfactory collector must avoid these two effects.

Fig. 11.3*a* illustrates one form of electrode means* for introducing the signal electrons into the multiplier which overcomes in a fairly satisfactory manner the difficulties mentioned. It consists of a disk, or plate, having a fairly large area, which is operated at a potential slightly positive with respect to the second anode. An aperture perforates the disk, and behind this aperture is mounted a short cone, or cylinder, which is at a high positive potential with respect to the disk. The electrons leaving the mosaic, which is approximately at disk potential, are prevented from reaching the second anode because of its relative negative potential. As they drift toward the disk they eventually enter the field from the cone and are guided toward and through the aperture.

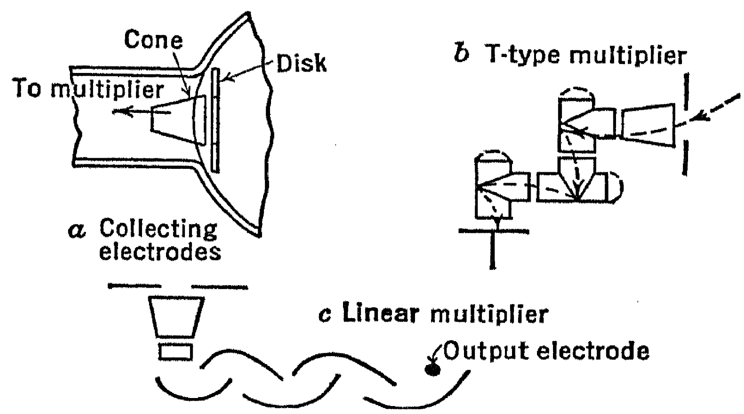


Fig. 11.3.—Secondary-Emission Signal Multipliers Used to Amplify Electron Current from Mosaic.

The multiplier is mounted behind the cone in such a way that the electrons coming through the cone strike the first target. A number of types of multiplier may be adapted for this arrangement. Broadly, they must satisfy the following qualifications: They should not require a magnetic field or a sharply focused electron beam on the first target, they should have a high gain per stage, and, finally, the activation schedule should not interfere with that necessary for the mosaic. Two multipliers which have been found satisfactory with the Iconoscope are illustrated in Figs. 11.3*b*† and 11.3*c*‡. Both are focused electrostatically, and neither requires a sharply defined initial beam of electrons. Of the two illustrated, the T-type multiplier permits a somewhat simpler geometric arrangement of tube but is more difficult to construct and activate, and also becomes space-charge-limited at a lower current value.

With the low maximum potential on the disk compatible with a uniform picture, the placement of the multiplier is important from the

* See Zworykin, Morton, and Flory, reference 1.

† See Zworykin, Morton, and Malter, reference 4.

‡ See Zworykin and Rajchman, reference 4.

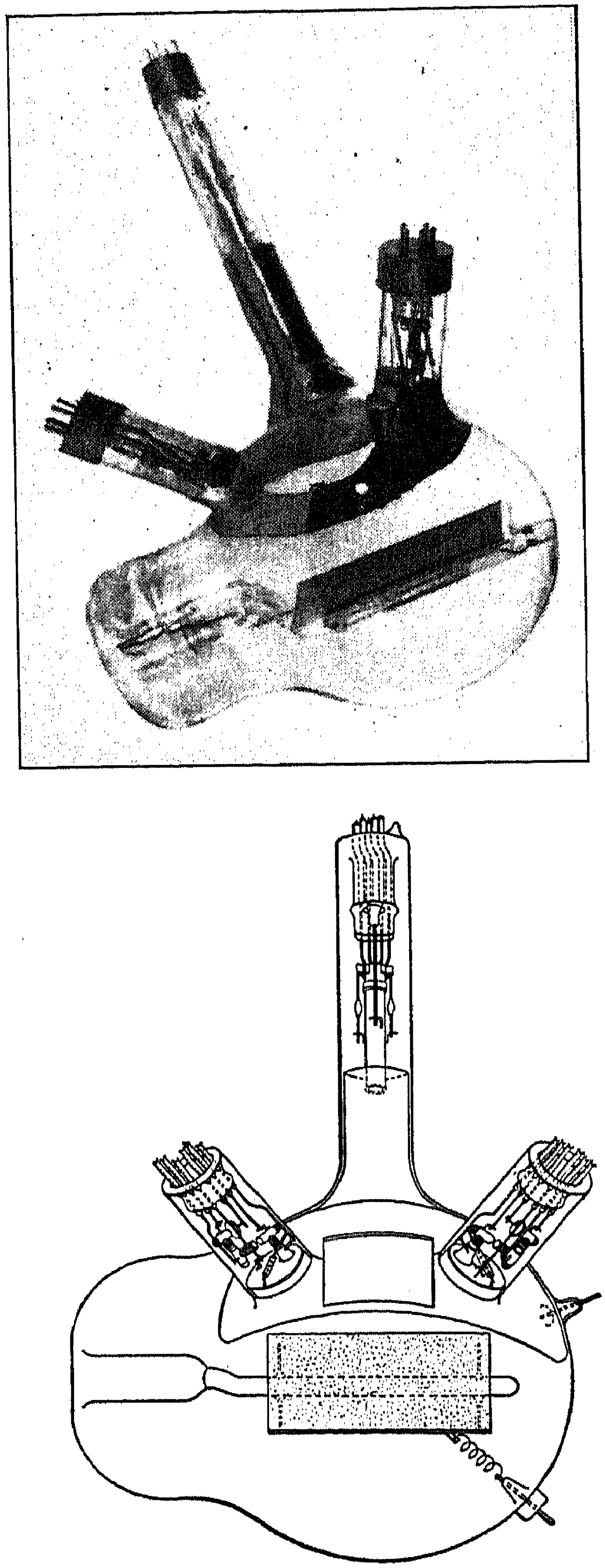


Fig. 11.4.—Signal-Multiplier Iconoscope.

standpoint of efficient collection. The following table gives the percentage of electrons leaving the mosaic collected by a multiplier variously located:

Location	Per Cent Collection
Behind mosaic	10
90° to normal	50
30° to normal	80
Normal to front of mosaic	100

For optical reasons, the multiplier cannot be located directly in front of the mosaic. However, by placing two multipliers at 30° on either side of the face normal of the mosaic, all the electrons leaving the mosaic can be collected, without interfering with the optical image.

A multiplier Iconoscope employing two T-type multipliers is shown in Fig. 11.4. Fairly satisfactory pictures have been obtained from laboratory models of this type of Iconoscope, at light levels considerably lower than are possible with the normal tube. However, many problems relating to uniformity of picture, efficiency, and saturated signal output still remain to be solved.

THE INCREASE OF AVAILABLE CHARGE

The third method of increasing the sensitivity of the Iconoscope, i.e., that of increasing the quantity of charge available per unit light flux, may be accomplished in a number of different ways, only one of which will be discussed in the present chapter.

11.9. The Image Iconoscope. The method to be considered is that of intensifying the image as a whole by secondary-emission multiplication.* In this type of tube the optical image is projected onto an extended photocathode. Electrons are emitted from this cathode with a density distribution identical with the intensity distribution of the light. Thus, close to the surface there will be an electron image which is an exact reproduction of the light image. By means of a suitable electron-optical system these emitted electrons are accelerated toward, and focused on, a mosaic. The mosaic thus has projected on it an electron image instead of the usual light image. The surface of the mosaic is activated in such a way that it is secondary-emissive, so that, for each photoelectron in the electron image which strikes it, several electrons are emitted. Bombarded portions of the mosaic, therefore, become positive with respect to their neighbors, just as do the illuminated areas of the mosaic in the normal Iconoscope. Consequently, when the mosaic is scanned, a picture signal corresponding to the light image on the photocathode is generated. A

* See Zworykin, Morton, and Flory, reference 1; Iams and Rose, reference 2; and articles listed in reference 5.

schematic diagram illustrating the principle of this type of tube is given in Fig. 11.5. Iconoscopes incorporating this principle have proved very successful and have been the subject of considerable investigation both in this country and abroad.

It is convenient, in the image Iconoscope, to use a semi-transparent photocathode. With this type of cathode the light image can be projected onto the rear surface and electrons will be emitted from its face. Since the photocathode is one of the most important elements in the image Iconoscope its preparation warrants further discussion.

A glass disk, often the end of the tube itself, may serve as a base for the cathode. This disk is arranged in the tube in such a way that a silver film can be evaporated on it during the activation treatment, after the tube has been assembled and evacuated. The source for evaporating the silver film, usually a tungsten filament on which silver beads have been melted, or a silver-coated tungsten helix, is mounted so that it does not interfere with the formation of the electron image. To meet this requirement, the evaporating filament may be mounted in an aperture of the electron lens system, or it may be made movable so that it can be withdrawn to a position where it does not affect the operation of the tube.

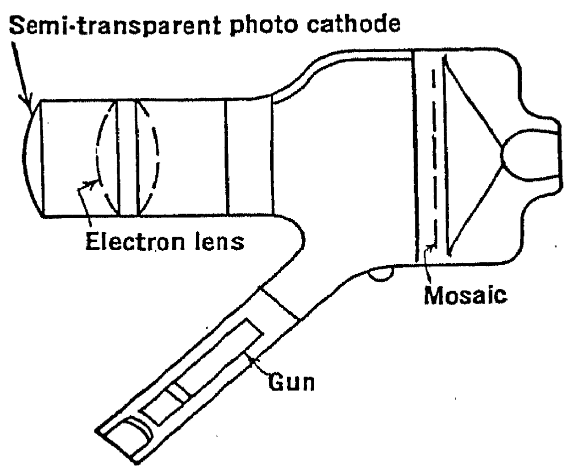


FIG. 11.5.—Schematic Diagram of Image Iconoscope.

After the tube has been evacuated and outgassed, a thin layer of silver is evaporated onto the cathode disk. This silver is then completely oxidized by a glow discharge in oxygen at a low pressure. When required the mosaic surface may be oxidized at the same time. Caesium is added from a side tube, as in the activation of the normal Iconoscope, and the tube is baked for a short period. The last step is silver sensitization, which consists of evaporating a small additional amount of silver and giving the tube a final bake.

This type of cathode will have a sensitivity of 20 to 30 microamperes per lumen. Films have been made which yield more than 50 microamperes, but the method has not been sufficiently perfected to make them usable for the image Iconoscopes.

Cathodes giving 30 microamperes per lumen in general have a color response identical with that of a hard caesium phototube, with its high response to red radiation. At a slight expense of sensitivity this red response can be reduced, which is desirable from the standpoint of fidelity

of image reproduction. The color responses of several typical films, made by processes similar to that outlined above, are shown in Fig. 11.6. Semi-transparent photocathodes employing evaporated bismuth, antimony, antimony, silver, etc., are also frequently used to obtain desired spectral response in image Iconoscopes.

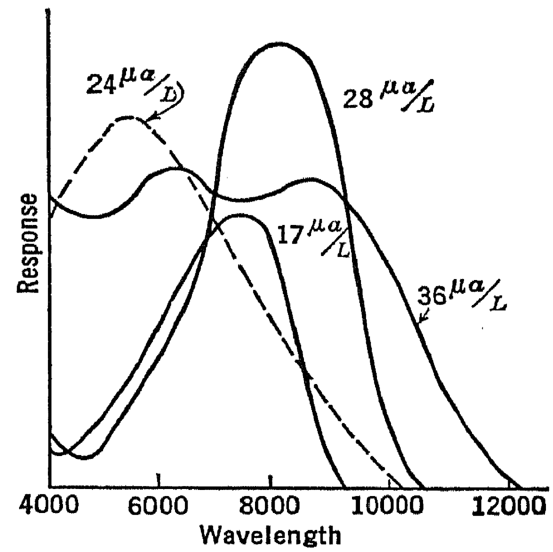


FIG. 11.6.—Spectral Response of Various Semi-Transparent Photocathodes.

A number of electron lens systems are applicable. Broadly, they may be divided into two classes: those which make use of a magnetic field as their principal focusing means, and those which are purely electrostatic.

A uniform magnetic field constitutes the simplest magnetic lens. Such a lens will produce an erect, distortion-free image whose principal aberration is that produced by initial velocities. For a number of reasons, this type of lens is not very well suited to the needs of the present image Iconoscope. Next to this in simplicity is the short-coil lens. This lens requires a magnetic field whose extent is small compared with the distance between object and image. The image formed by the short coil will be inverted and will have a magnification given by the ratio of the image-to-coil distance and the object-to-coil distance. In practice the short-coil conditions cannot be truly fulfilled, unless the diameter of the coil is comparable with that of the electron ray bundle, which results in serious aberrations. Experimentally, a lens which lies between these two extremes has been found to be most suitable. This lens consists of a coil which extends over most of the electron path between cathode and mosaic. In order to minimize aberrations, it has been found that the magnetic flux lines near the cathode should be parallel to the trajectories that would have been assumed by the electrons had the magnetic field not been present. Such a field can be obtained with a non-uniform coil having its thickest portion at the cathode end, as shown in Fig. 11.7. A lens of this type produces a real image which is rotated by about 30°. The electrons, in traversing the distance between cathode and mosaic, do not cross the axis of the lens system but execute a helical path such that their

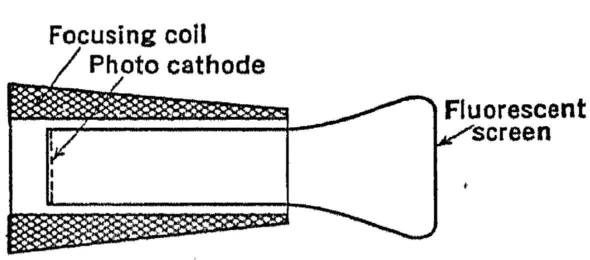


FIG. 11.7.—Image Tube with Magnetic Lens.

radial distance changes only by the amount required to give the image magnification. The difference in convergence of ray bundles originating from various radial distances on the cathode is less than for systems in which the electrons cross the axis, and, consequently, the curvature of the image surface is less.

The most important image defects in this system are:

Curvature of image field.
Pincushion distortion.
Rotational distortion.
Chromatic aberration.

The first two are sufficiently small that it is not necessary to correct for them further. Rotational distortion can be minimized by properly choos-

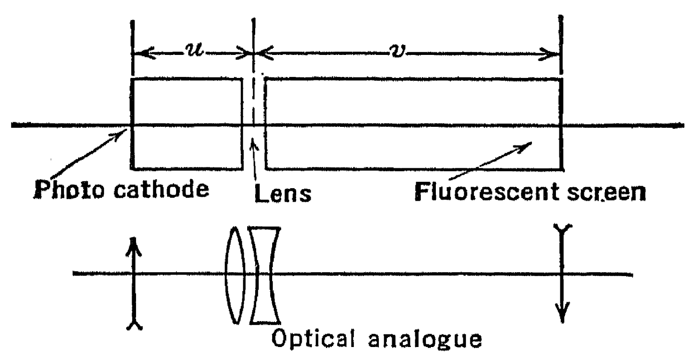


FIG. 11.8.—Simple Electrostatic Image Tube.

ing the length, position, and distribution of turns of the lens coil. Chromatic aberration, i.e., the aberration caused by the spread of initial velocities of the photoelectrons, cannot be eliminated, but, by the application of fairly high overall voltages, it can be reduced to a point where it does not limit the system.

From a theoretical point of view an electrostatic lens system is much simpler than one using a magnetic lens. Structurally, they present a problem of equal magnitude.

One of the simplest electrostatic lenses which can be employed for the required imaging is that formed by two coaxial cylinders, as illustrated in Fig. 11.8. Such a lens system was discussed in Chapter 4. The cathode closes the end of one of the two cylinders; the screen is at the opposite end of the other. When a potential is applied between the cathode and anode cylinders, the latter being positive, electrons leaving any point on the cathode will converge onto a point in the anode side of the system, this point constituting the image of the original point of emission. As a matter of convenience, the plane between the ends of the cathode and anode cylinders will be called the lens, although actually the lens action

extends over most of the electron trajectory. The relation between the cathode-to-lens distance, u, and the lens-to-image distance, v, is given in Fig. 11.9a, while the magnification as a function of these parameters is shown in Fig. 11.9b. For a given distance between the cathode and lens, and a given cylinder diameter, the position of the electron image is fixed.

From a constructional point of view, it is highly desirable that the lens be made to focus electrically. This can be done by separating the cathode from the cathode cylinder, and applying a potential to the latter to strengthen or weaken the convergence of the lens. If, however, the dimensions of the system are not close to those required for fixed focus, a considerable difference in potential between the cathode and cathode cylinder is required for focusing, which seriously distorts the image. To

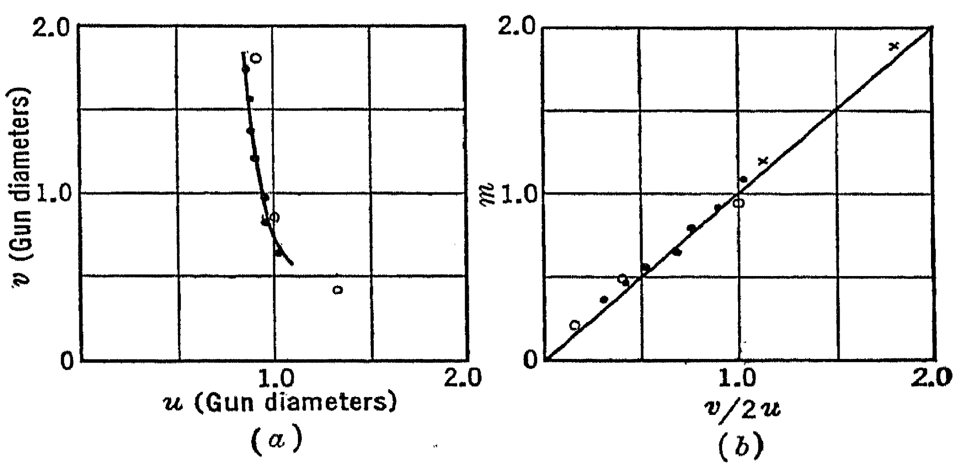


Fig. 11.9.—Focusing Properties of Electrostatic Image Tube.

avoid this distortion, the cathode cylinder can be made of a resistive material so that the cathode end can be maintained at cathode potential, while the other is raised to some focusing potential. In practice, the cathode cylinder is divided into a series of four or five rings, and the potential difference between the cathode and the cathode side of the lens is divided between them.

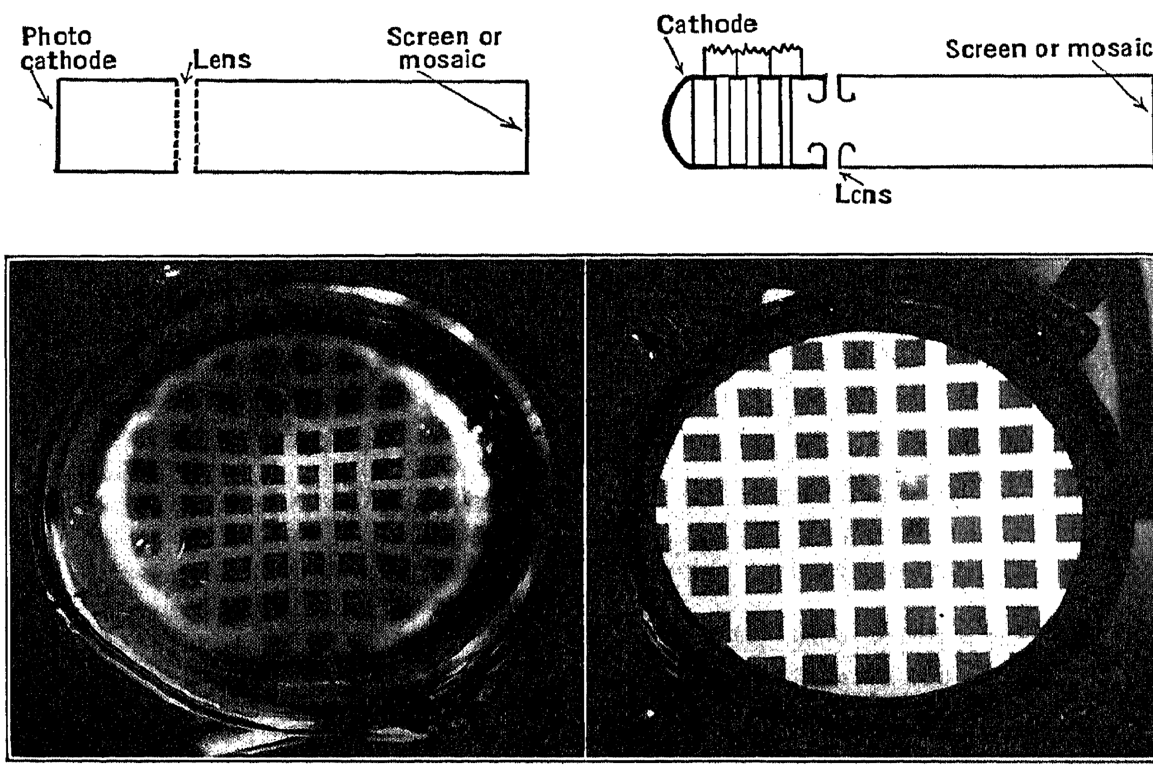
The aberrations in this type of system have been studied in considerable detail, both theoretically and experimentally. The most serious image defects are as follows:

Curvature of image field.
Pincushion distortion of the image.
Astigmatism.
Chromatic aberration.

Spherical aberration and coma are negligible in this type of system, as was pointed out in Chapter 4.

By properly shaping the cathode the first three defects can be largely overcome. Experimentally, it has been found that, for a self-focusing

system, a cathode having a radius of curvature equal to the lens diameter reduces these aberrations to such an extent that they do not limit the system (Fig. 11.10*b*). A comparison of Fig. 11.10*a* with Fig. 11.10*b* shows the extent of this correction. Unless very elaborate correcting systems are resorted to, the only means of reducing chromatic aberration is to use a high overall potential, thus reducing the ratio of initial velocity to working velocity. Tests indicate that, with 1000 volts applied to the system, about 300 lines can be resolved, while at 2000 volts, a resolution of 500 lines or better can be expected. At higher voltages,



a *b*

FIG. 11.10.—Correction of Image Defects by Curvature of the Cathode.

mechanical and electrical imperfections of the cathode and lens usually set the limit to the resolving power.

It should be pointed out that the curved cathode, though making it possible to obtain an undistorted electron image, introduces a serious optical problem. An objective with a relatively small aperture will have sufficient depth of focus to take care of the curvature of the image required to fit the cathode. However, when the numerical aperture is large, it is necessary to modify the optical system used with this electron lens.

The mosaics for the image Iconoscope may be broadly divided into two classes: namely, two-sided and single-sided mosaics.

A two-sided screen would be very desirable in that the secondary

emission from the image electrons could be saturated, giving the full benefit of the image intensification principle. However, the same difficulties are encountered in the construction of a two-sided mosaic for this purpose as were mentioned for the photoelectric two-sided mosaic.

Very excellent results have been obtained with the image Iconoscope using a single-sided mosaic. A mosaic for this purpose must meet the following requirements:

1. The capacity between the sensitized surface and the signal plate should be about 100 micro-microfarads per square centimeter. This value is not at all critical, but must be uniform over the entire mosaic.
2. The resistance must be high, both through the dielectric and over its surface.
3. The secondary-emission ratio and initial velocities must be high.

A large number of mosaics are found suitable. Three common types will be described.

The first is very similar to the photosensitive mosaic in the normal Iconoscope. One side of a freshly cleaved sheet of mica 0.001 to 0.002 inch thick is coated with a conducting signal plate, and the other with a vast number of silver elements deposited as described in the preceding chapter. It will be found that a more satisfactory mosaic is obtained if the silver coverage is somewhat lower than for the photosensitive surface. After it has been mounted and the tube exhausted, the silver surface is oxidized and activated with caesium. One thing that recommends this type of surface is that it is comparatively easy to obtain a uniform surface free from blemishes.

A second mosaic, which leads to greater sensitivity and permits, in general, higher resolution, can be made as follows. A sheet of mica, of the same thickness as for the previous mosaic, is coated on the back with a signal plate, but its face is left free. The surface is activated by glowing in an oxygen discharge and exposing it to caesium vapor. Both the secondary-emission ratio and the sensitivity are higher than for the silver mosaic. However, it is more difficult to obtain uniformity with mica activated in this way.

The third form of mosaic differs from those previously described in that, instead of mica as the dielectric, a finely divided insulating powder is used. The powder is applied as a coating on a metal sheet which serves as the signal plate. The response for this type of mosaic is considerably higher than can be obtained with either of the others. With a proper technique for handling the powder, very excellent mosaics of this type can be obtained, with respect not only to sensitivity but also to uniformity.

Typical secondary emission and response curves for these mosaics are shown in Fig. 11.11.

The operation of the single-sided mosaic in the image Iconoscope is identical with that in the normal Iconoscope, the secondary emission from the image electrons playing the same role as the photoemission in the normal Iconoscope. It is, therefore, unnecessary to discuss the theory at this point. Mention should be made, however, of the fact that, since the emission is not saturated, the initial velocities of the secondary electrons, as well as the ratio, determine the final sensitivity of the tube.

The final element, the electron gun, is of purely conventional design, and will be described in detail elsewhere. The beam voltage used with this type of tube is in the neighborhood of 1000 volts, and the current in

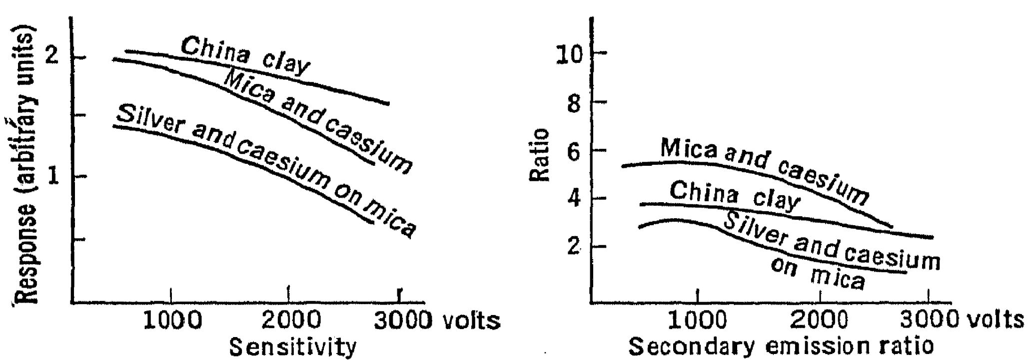


FIG. 11.11.—Properties of Various Mosaics Suitable for the Image Iconoscope.

the scanning beam is between 0.1 and 0.2 microampere. In this respect, it resembles the normal Iconoscope.

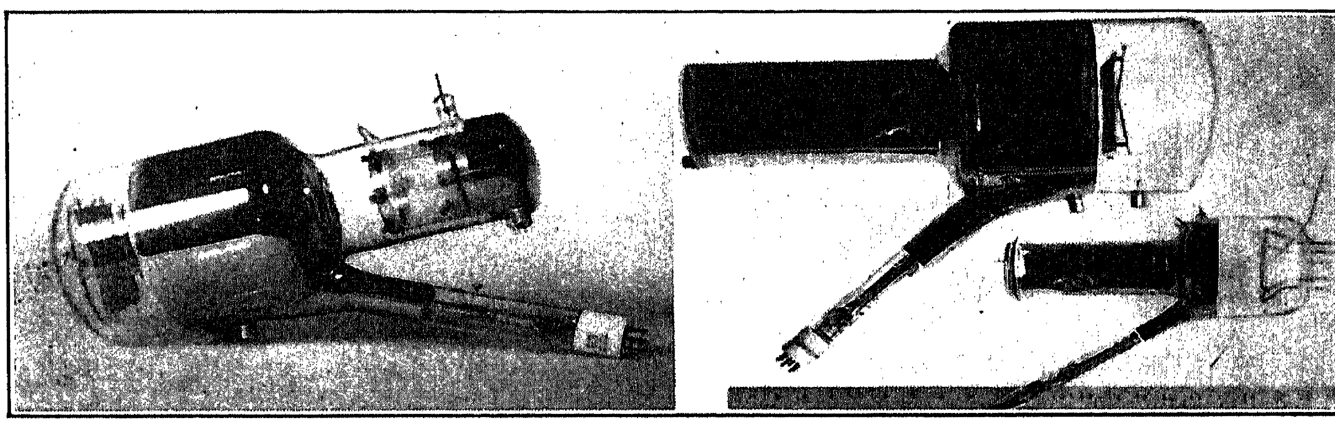
The advantages gained in this tube wherein an electron image is projected onto a mosaic are as follows:

It is possible to make semi-transparent photocathodes which have a photosensitivity of 20 to 30 microamperes per lumen. This is to be compared with a maximum of about 15 microamperes per lumen for the mosaic of the normal Iconoscope. Even more important is the fact that a strong field exists at the photocathode which saturates the emission. When it is remembered that the efficiency of a mosaic as a photocathode is only 20 to 30 per cent, owing to the unfavorable conditions, the advantage of this method as far as the photoelectric emission is concerned is very apparent.

The photoelectrons are focused onto the mosaic, which has a high secondary-emission ratio. The actual measured ratios for various types of mosaic range from 3 to 11. However, since the same unfavorable field conditions are found at the mosaic as exist in the normal Iconoscope, the secondary emission will not be saturated and, consequently, only a fraction of the maximum emission will be available for forming the charge

image. The magnitude of the fraction available will depend upon the distribution of initial velocities of the secondary electrons from the substance in question. If there is a large percentage of high-velocity electrons, the fraction of the total secondary-emission ratio available to form the charge image will be greater, and for a given ratio both saturated signal output and the sensitivity at low light levels will be higher. The relation here is very complicated, however, since these same factors influence the behavior of the electrons produced by the beam and, consequently, both the redistribution and the potential distribution on the mosaic.

Two RCA image Iconoscopes, developed to a point where they are adequate for practical transmission, are shown in Figs. 11.12*a* and 11.12*b*.



a *b*

Fig. 11.12.—Electrostatic and Magnetic Image Iconoscopes.

That illustrated in Fig. 11.12*a* is electrostatically focussed, necessitating a curved cathode. This tube is operated with the cathode between 2000 and 3000 volts negative with respect to ground. The focusing voltage can be supplied from a potential divider between cathode and second anode, thus minimizing the filtering required for the voltage source. Fig. 11.12*b* shows two magnetically focused tubes, without lens coils. This type of tube when operated at about 1000 volts requires a flux density of 140 gauss for the lens field.

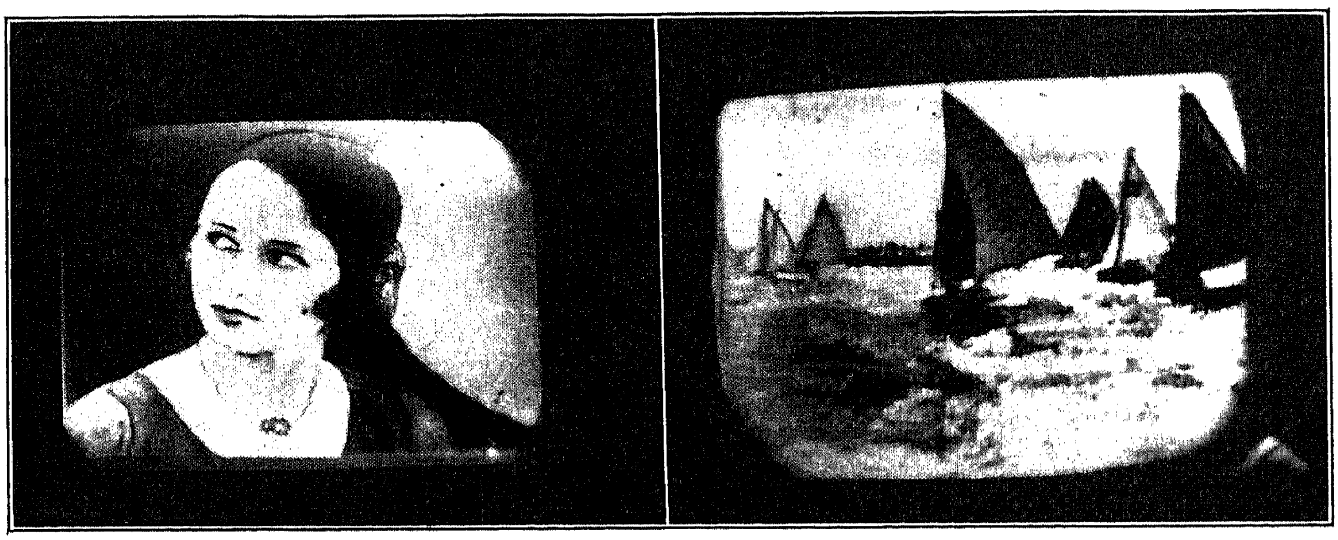
Although the operating conditions of the image Iconoscope are similar to those of the normal Iconoscope, there are differences which are important and should be noted.

Because of the use of an electron image tube, there are three focusing adjustments to be made: namely, that of the optical image, the electron reproduction, and the scanning beam. In actual operation, the last two are adjusted once and for all, and only the optical focus need be changed to meet the requirements of different scenes.

The deflecting field of the scanning beam tends to disturb the electron

image, particularly when the tube is operated at a low voltage. Suitable shielding around the deflecting yoke completely eliminates this interference.

The earth's magnetic field constitutes another source of interference. The magnetic vector of the earth's field at a latitude of 42°N makes an angle of about 70°, so that its effect is to displace the image to one side and either to shift it up or down or to produce a slight rotation, depending upon the direction of the axis of the tube. There are, obviously, a number of ways of correcting for this, as by shielding, compensating coils, etc.



a *b*

Fig. 11.13.—Pictures Obtained with an Electrostatic and a Magnetic Image Iconoscope.

The optical image on the photoelectric cathode in this type of tube is much smaller than that on a normal mosaic, and, if the angular field of the objective is to be the same as for the normal Iconoscope, the focal length of the lens must be proportionately smaller. In order that such a lens have the same total light-gathering power and depth of focus in the image field, its numerical aperture is, of course, larger than that of the longer-focal-length lens.

The two photographs reproduced in Figs. 11.13*a* and 11.13*b* were made from scenes picked up by the image Iconoscope. The quality of the picture is nearly as high as that from the normal Iconoscope, in respect to both resolution and freedom from imperfection, and the light required when an equivalent lens is used is between 1/6 and 1/10 as great.

11.10. Multi-Stage Image-Multiplier Iconoscopes. Up to this point the discussion has been limited to image Iconoscopes having only one stage of image intensification. Experimental Iconoscopes have been made and operated in which the electron image is projected onto a secondary

emitting surface, and the emitted electrons refocused onto a secondary emissive mosaic. Although the development of the multi-stage tube is a long way from the practical perfection of the single-stage tube, nevertheless it is a very promising field.

It is interesting to estimate the limiting sensitivity of a multi-stage image Iconoscope, wherein the number of stages is sufficient so that the statistical fluctuations of the electrons from the photocathode determine this limit. In order to make such an estimate, certain simplifying assumptions are necessary whose correctness cannot, at present, be verified, so that the conclusions arrived at here may be radically altered as the investigation of this type of tube proceeds.

The photocathode can be imagined to be divided into m small areas corresponding to picture elements. If the mean charge emitted by one such element during a frame time is q, then the mean square fluctuation of charge, $\overline{q_N{}^2}$, will be given by:

$$\overline{q_N{}^2} = eq. \tag{11.7}$$

The emitted charge, q, is multiplied by the image intensifier and reaches the mosaic as a charge Gq.

The signal-to-noise ratio, referred to the individual element, which is consistent with the previous method of noise calculation, can be written as:

$$S = \frac{q}{\sqrt{\overline{q_N{}^2}}} = \frac{q}{\sqrt{eq}} = \sqrt{\frac{q}{e}}. \tag{11.8}$$

In this case, since q is the charge accumulated on an element in a picture time, it is given by:

$$q = i_e T_0 = \frac{LAp}{m} T_0.$$

Therefore:

$$S = \sqrt{\frac{LAp}{em/T_0}}.$$

It can be shown by means of a Fourier transformation that m/T_0 can, in this equation, be replaced by $2F$, where F is the frequency bandwidth required to resolve the m elements. Thus, the ratio becomes:

$$S = \sqrt{\frac{LAp}{2eF}}. \tag{11.9}$$

Actually, the signal-to-noise ratio cannot be as great as this because the multiplied photoelectrons impinging on the secondary-emissive mosaic

introduce an additional noise which, since it has the same form as that just calculated, can be introduced into Eq. 11.9 as a factor $(1 + k'')$, where k'' is the additional noise which must always have a value greater than unity. The final form of the signal-to-noise ratio is:

$$S = \frac{1}{1 + k''}\sqrt{\frac{LAp}{2eF}}.$$

It is interesting to compare this with the limiting noise for the low-capacity signal multiplier Iconoscope, which may be written as:

$$S = \sqrt{\frac{LAp}{2eF}\frac{k^2(B - 1)}{k'(2B - 1)}}.$$

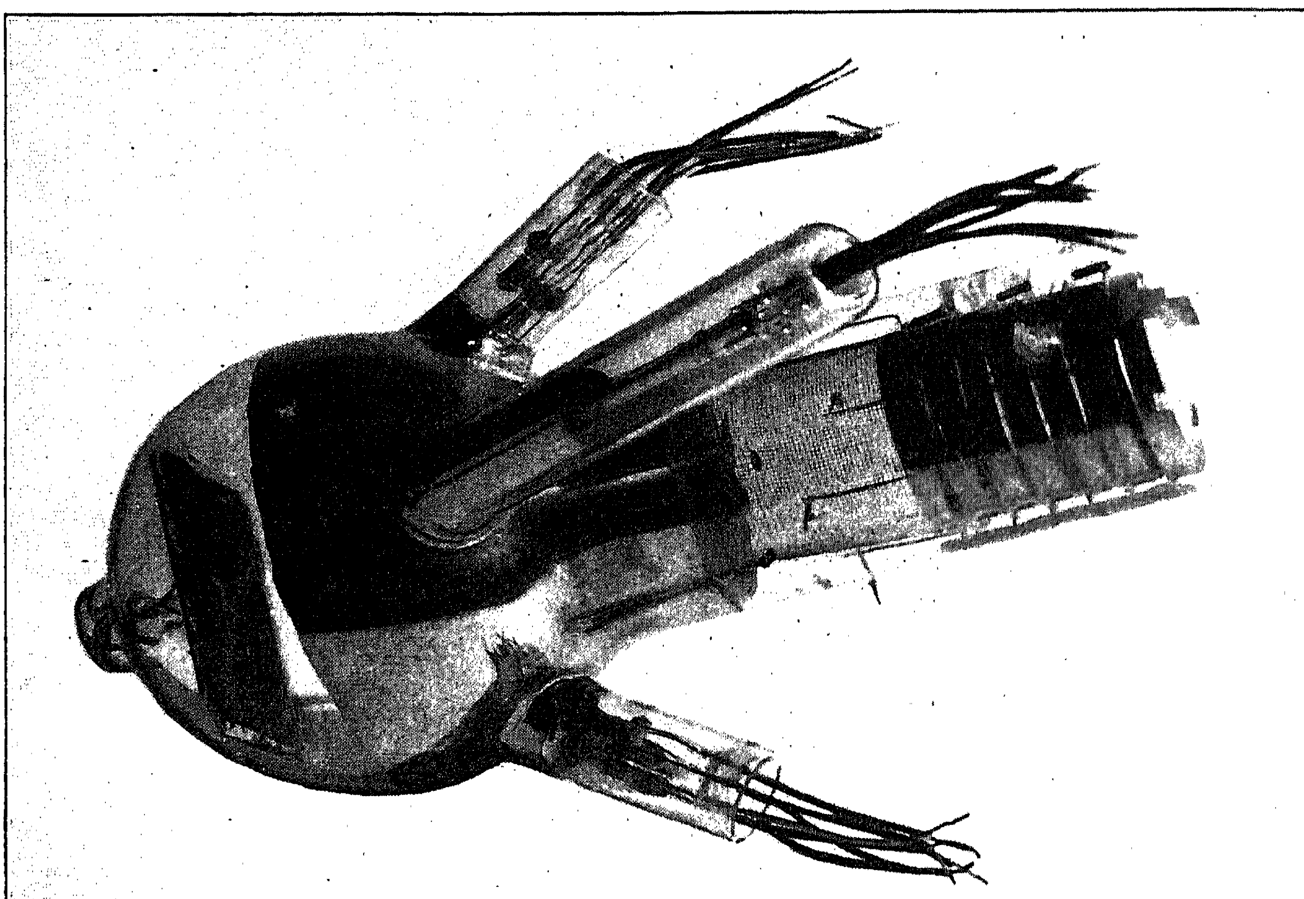


FIG. 11.14.—Image Iconoscope with Signal Multipliers.

The form is the same and the equations differ only by the efficiency factors, $1/(1 + k'')$ and

$$\sqrt{(k^2(B - 1))/(k'(2B - 1))}.$$

Because it has not yet been possible to realize the theoretical maximum sensitivity of either the image intensifier or signal multiplier Iconoscope,

an advantage can be effected by combining the two principles in a single tube. A tube of this type is illustrated in Fig. 11.14, employing two multipliers and a powder insulated mosaic.

SPECIAL TUBES

There are a great many special tubes for generating a picture signal through the action of a scanning beam sweeping across a suitably prepared target. Most of these tubes derive their signal through a change in the secondary-electron current produced by the beam. The change in current may be the result of the action of light, heat, or other radiation on the target, a permanent pattern marked on the screen, or charges electrically induced on the surface of a special mosaic.

11.11. Photoconductive-Screen Tubes. Certain materials are known to change their resistance under the action of light.* Examples of such substances are selenium, zirconium oxide, and aluminum oxide. When a conducting screen is coated with one of these photoconductive materials, the resistance variation over its surface will correspond to the variation of light in any image projected thereon. If, now, such a screen is scanned, the back plate being made negative with respect to the second anode by an amount less than the saturation voltage of the electrons leaving the screen, a picture signal can be obtained from a light image. The mechanism of signal generation can best be explained with reference to the curves shown in Fig. 11.15. The curve AB represents the saturation curve of current reaching the second anode as a function of the voltage between the surface of the screen and the second anode. The back plate is maintained at a potential represented by point E. The current and voltage conduction curves for an illuminated and an unilluminated area are given by lines M and N. Since the resistance decreases with light and the slope of the conduction curve is inversely proportional to the resistance, an increase in light represents an increase in slope. The intersection of the saturation curve and the conduction curve is the operating point of the surface. From the figure it will be seen that the current reaching

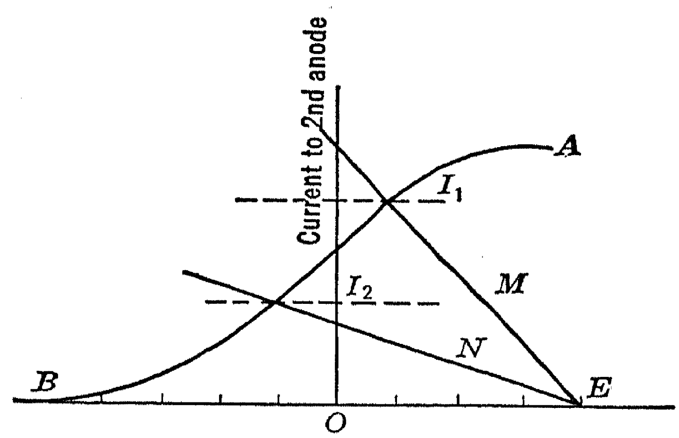


FIG. 11.15.—Current Characteristic of Photoconductive Screen.

* See Iams and Rose, reference 2; Knoll and Schröter, reference 6; and von Ardenne, reference 7.

the second anode changes from I_1 to I_2 as the beam moves from an unilluminated region onto one that is illuminated.

This type of tube does not employ the storage principle directly. However, most photoconductive phenomena are cumulative in nature, which introduces a form of storage. In fact, this ability to accumulate and retain the change of resistance is so great, in a number of materials, that it causes the reproduced image to lag behind changes in the light image for a considerable length of time.

11.12. Photovoltaic Screens. Another well-known form of photoactivity can be used as a basis for pickup tubes. This phenomenon is the ability of some substances, e.g., copper oxide, to develop a potential under the action of light. Under suitable operating conditions a picture signal can be obtained from a screen coated with a photovoltaic material. The signal is the result of the change in secondary-emission current consequent on the variation in potential over the surface due to difference in illumination. To obtain the maximum current variation the screen should

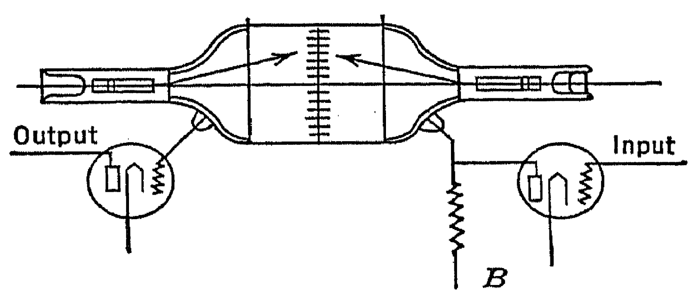


FIG. 11.16.—Storage Tube.

be biased to a point on the secondary-emission saturation curve where the ratio is varying rapidly with screen voltage. It should be pointed out that the action will not be the same for a high-resistance film which is driven to an equilibrium potential by the beam, and a low-resistance surface which remains at the potential established by the illumination. Photovoltaic surfaces are often photoconductive as well, and, when they are used in a pickup tube, the signal will be the result of the two mechanisms. The surface must, in this case, be arranged so that the effects aid rather than oppose one another.

11.13. The Storage Tube. An interesting application of a mosaic consisting of insulated elements is found in the storage tube. This tube may take any one of many forms. The description of one will illustrate the class. Fig. 11.16 shows this tube diagrammatically. It consists of a two-sided mosaic which is scanned on either side by separate guns. The signal to be stored is applied to one of the two second anodes. As the beam sweeps across this side of the mosaic, the elements will assume a potential

which is dependent upon the instantaneous potential of the second anode. When the scanning pattern is complete, a potential variation exists over the mosaic which is a record of the voltage variations applied to the second anode. For example, a potential image may be thus recorded on the mosaic. When scanned from the other side, this record is transformed into a signal collected from the other second anode. Such a tube can be used where it is desired to delay a picture signal by any amount up to one scanning period.

11.14. The Monoscope. The Monoscope is another very useful, special tube.* This tube resembles an Iconoscope in make-up, but in place of the mosaic it has a target upon which is printed the image whose video signal it is desired to reproduce. The back plate is made of a metal having

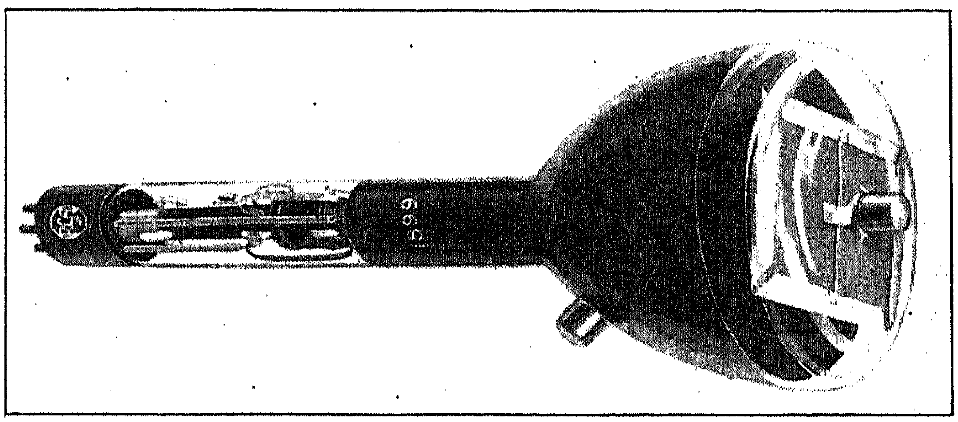


FIG. 11.17.—Monoscope RCA 1899.

a fairly high secondary-emission ratio, for example, aluminum, which has an oxide coating due to atmospheric oxidation. The pattern to be reproduced is printed on this surface in carbon ink, or other material having a different secondary-emission ratio from the back plate. When the pattern is scanned, the secondary-emission current collected by the second anode will vary in such a way as to give a picture signal of the printed pattern. This signal may be obtained either from the second anode or from the back plate. If the signal is taken from the back plate and it is desired that the video signal be in the same sense as that from an Iconoscope, a negative of the pattern must be printed on the back plate. A typical Monoscope is illustrated in Fig. 11.17.

This type of tube is very valuable as a means of generating a video signal for testing amplifiers, deflecting systems, transmitters, and viewing tubes.

The discussion of special tubes just concluded is not intended to be exhaustive but merely to indicate a few of the vast number of possibilities which exist in this field.

* See Burnett, reference 8; and Knoll, reference 9.

REFERENCES

1. V. K. Zworykin, G. A. Morton, and L. E. Flory, "Theory and Performance of the Iconoscope," *Proc. I. R. E.*, Vol. 25, pp. 1071–92, August, 1937.
2. H. Iams and A. Rose, "Television Pick-up Tubes with Cathode Ray Beams," *Proc. I. R. E.*, Vol. 25, pp. 1048–70, August, 1937.
3. A. Rose and H. Iams, "Television Pickup Tubes Using Low-Velocity Electron-Beam Scanning," *Proc. I. R. E.*, Vol. 27, pp. 547–555, September, 1939.
4. V. K. Zworykin, G. A. Morton, and L. Malter, "The Secondary Emission Multiplier—A New Electronic Device," *Proc. I. R. E.*, Vol. 24, pp. 351–375, March, 1936.
   W. Shockley and J. R. Pierce, "The Theory of Noise for Electron Multipliers," *Proc. I. R. E.*, Vol. 26, pp. 321–332, March, 1938.
   V. K. Zworykin and J. A. Rajchman, "An Electrostatic Electron Multiplier," *Proc. I. R. E.*, Vol. 27, pp. 558–566, September, 1939.
5. H. Iams, G. A. Morton, and V. K. Zworykin, "The Image Iconoscope," *Proc. I. R. E.*, Vol. 27, pp. 541–547, September, 1939.
   Nagashima, Shenozaki, Udagawa, and Kizurka, "The 'Tecoscope' C-R Television Transmitter," *Rept. Radio Research Japan*, Vol. 7, pp. 12–13, June, 1937.
   "Super Emitron Camera," *Wireless World*, Vol. 41, pp. 497–498, Nov. 18, 1937.
   J. D. McGee and H. G. Lubszynski, "EMI Cathode-ray Television Tubes," *J. I. E. E.*, Vol. 84, pp. 468–475, April, 1937.
6. M. Knoll and F. Schröter, "Translation of Electron Pictures and Drawings with Insulating and Semi-Conducting Layers," *Physik. Z.*, Vol. 38, pp. 330–333, May, 1937.
7. M. von Ardenne, "On Experiments with Photosensitive Semi-Conducting Layers in Cathode Ray Tubes," *Hochfrequenztechn. u. Elektroakustik*, Vol. 50, pp. 145–149, February, 1938.
8. C. E. Burnett, "The Monoscope," *R. C. A. Rev.*, Vol. 2, pp. 414–420, April, 1938.
9. M. Knoll, "Electron Optics in Television Technique," *Z. tech. Physik*, Vol. 17, pp. 605–617, 1936.
10. G. Krawinkel, W. Kronjäger, and H. Salow, "On the Question of Electrical Picture Storage," *Telegraphen, Fernsprech, Funk and Fernseh-Technik*, Vol. 27, pp. 527–533, November, 1938.
11. G. Krawinkel, W. Kronjäger, and H. Salow, "On a Storing Picture Pick-up with Semi-Conducting Dielectric," *Z. tech. Physik*, Vol. 19, pp. 63–73, March, 1938.

CHAPTER 12

THE KINESCOPE

The Kinescope, as was indicated in Chapter 9, is one of the two terminal tubes of the television system. The picture, as a consequence of a long chain of physical events, is reproduced in light and dark on its viewing screen. The position of the viewing tube as the final link in the television chain places upon it the responsibility of converting a mass of electrical phenomena into a visible image allowing accurate cognizance of the scene being transmitted.

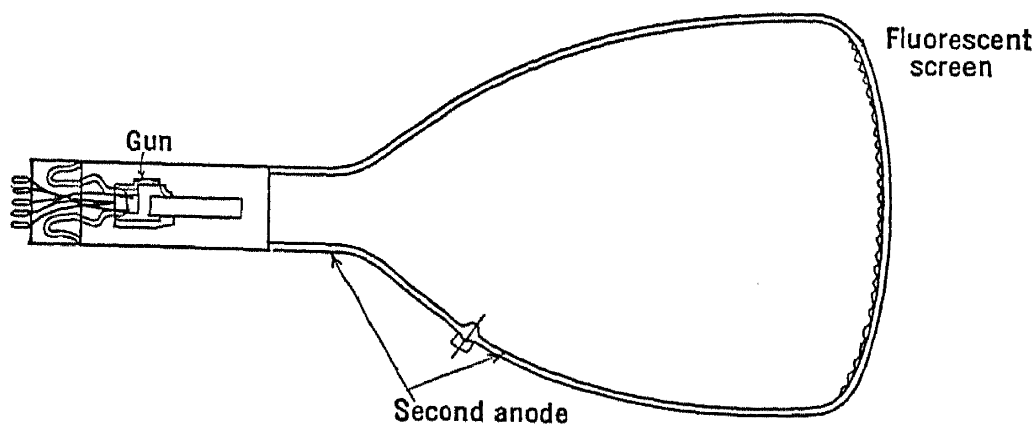


FIG. 12.1.—Diagram of Kinescope.

Basically, the Kinescope is quite simple, consisting of an electron gun and fluorescent screen assembled in a highly evacuated bulb. The electron gun, like that in the Iconoscope, is an electron-optical system which concentrates the electrons from a thermionic cathode into a narrow pencil. It is located in a narrow neck which terminates the pear-shaped main body of the tube. Opposite the gun is the slightly curved broad end of the bulb, and on this surface is coated the fluorescent material or phosphor. The arrangement of the elements can be seen from Fig. 12.1.

The cathode-ray pencil from the gun strikes the screen in a small area called the "spot." Owing to the luminescent properties of the screen the "spot" is visible as a bright point of light. The current in the electron beam and, therefore, the amount of light emitted at the spot can be varied by a potential applied to a control element in the gun.

Means are provided, in the form of either electrostatic deflecting plates or magnetic coils, for deflecting the spot to any position on the screen.

The spot can therefore be made to sweep across the Kinescope screen in the series of straight, parallel lines required for the scanning pattern by means of suitably varying potentials or currents applied to these deflecting means. The entire scanning pattern is swept out at such a rate that, by virtue of the persistence of vision, the pattern is seen as static. That is, when examined closely the screen of the Kinescope appears to be ruled with a grating of fine luminous lines, but at a distance of a few feet it appears uniformly luminous.

At the pickup end of the system an exploring spot sweeps out a similar pattern. Not only are the scanning patterns similar in number of lines and in aspect ratio, but also the spot at the receiver is so synchronized with that at the transmitter that the two spots are at every instant at the same relative positions on their respective patterns. Thus, if the brightness of the spot at the receiving end is controlled by the amplitude of the video signal, a reproduction of the image at the transmitter will be formed on the viewing screen. By supplying the video signal to the control grid of the Kinescope gun this variation in spot brightness can be obtained and a luminous reproduction of the picture being televised will be formed on the fluorescent screen.

The generation of the video signal has already been discussed in Chapters 10 and 11. The amplification, transmission, and reception of the video signal, prior to applying it to the grid of the Kinescope, will be taken up later. Also the very important problem of synchronizing and deflecting the beam must be reserved for another chapter. The sole concern of the present chapter is, therefore, the requirements, construction, and performance of the tube on which the picture appears.

12.1. Requirements of the Kinescope. The general requirements of a high-definition picture have already been discussed without reference to any particular television system. When referring to a particular means for reproducing a picture, some of these requirements must be tempered because of practical considerations, and certain compromises must be made.

The size of the picture reproduced by a direct-viewing Kinescope determines the size of the tube. Various tests have indicated that, for general family use in the home, a picture between 1 and 2 feet on the diagonal would be desirable. A picture which is larger than this has certain advantages, but the general feeling is that a picture which is much larger could not easily be fitted into the average home decoration or environment. Considering the various factors such as production problems, cost, physical size of equipment, and picture requirement, tubes having screens 12 inches and 15 inches in diameter have been chosen as most desirable for ordinary console television receivers. From the standpoint

of tube construction this range of size can be readily realized. Smaller tubes will also have their place in sets for which the price is an important consideration. Very much larger pictures will undoubtedly be required for public entertainment. The projection Kinescope and large demountable tubes, which are capable of performing this function, will be considered at the end of the chapter.

Psychologically it would be desirable to have the picture appear on a flat screen. However, where the end of the tube serves as the viewing screen the very great reduction in thickness of glass that can be effected by slightly curving the end of the tube more than offsets the advantage of a flat screen.

One of the most important considerations in connection with picture quality is resolution. It was shown in Chapter 6 that a picture which is made up of 400 to 500 horizontal lines meets, at least for the present, the requirements of high-definition television. Assuming video signals capable of giving this resolution the factors necessary in the Kinescope are a small scanning spot, accurate deflection, and fine grain texture of fluorescent screen material. For a picture about 12 inches on the diagonal and a 4:3 aspect ratio (i.e., 10 by 7½ inches) the spot diameter must be less than 0.02 inch. The size of the particles of screen material should be such that at least four grains are under the beam at any one instant, and preferably many more. The deflection must be accurate to a fraction of the separation between lines.

The brightness of the screen is an important factor in the entertainment value of the picture, particularly if it is to be viewed over a long period of time. The brightness required is not a fixed quantity, but depends upon the average illumination of the surroundings and, to some extent, upon the content of the picture being viewed. If the receiver is in complete darkness, the light level required to avoid undue fatigue will be much lower than if it is in a moderately lighted room. This is so, of course, because the accommodation of the eye is determined by the average illumination entering the pupil. Similarly, a black-and-white cartoon can be viewed without fatigue with less light than is required for a half-tone picture. The brightness of a television screen must be sufficient so that an ordinary half-tone reproduction can be viewed with comfort in a normally lighted room.

For orientation Table 12.1 lists the brightness of a few familiar objects.

TABLE 12.1

	Foot-Lamberts
Outdoor scene (bright day)	300 to 600
Moving picture theater screen (high light)	2.7 to 5.2
Lighted page (minimum recommended)	10

From such data and various tests it has been concluded that the brightness of the screen must be 10 or more foot-lamberts—a brightness which can easily be obtained on a Kinescope screen. The major factors governing the light which can be obtained from the screen are the fluorescent material itself, the current in the beam, and the voltage of the bombarding electrons. To a much less extent it is a function of spot size (due to saturation) and scanning frequency (due to phosphorescent decay time).

The color of the luminescence of the screen is, in a sense, related to the brightness in that the eye has a varying response over the visible spectrum. Green, or yellow-green, not only produces the maximum sensation for a given amount of light, but also is least fatiguing to watch. These reasons make it a desirable screen color. However, tests over a long period of time have revealed that these advantages are more than offset by the adverse psychological effect of any color other than white. Therefore, white or nearly white screens have come into wide usage.

The scanning process, when referring to any small area of the screen, is essentially a discontinuous one and, as such, introduces the possibility of flicker. This flicker, when pronounced, is extremely objectionable. The magnitude of the sensation of flicker depends upon vertical sweep frequency, duration of phosphorescence, and brightness of the screen. Irrespective of the second two factors, flicker can be made imperceptible if the vertical sweep frequency is sufficiently great. Increase in the time of phosphorescent decay decreases the threshold frequency; increased brightness has the opposite effect, as has heretofore been pointed out. With the fluorescent materials available and the brightness ordinarily used, the flicker threshold occurs at a much higher frequency than that necessary for continuity of motion. In order to avoid the impression of flicker without necessitating the excessive bandwidth which would be the result of a repetition rate which is higher than that required for continuity, interlaced scanning is used.

For the mere conveyance of information, contrast is relatively unimportant. However, it is a factor of major importance in determining the quality of a picture. The meaning of the term contrast and its relation to noise, brightness, and visual threshold have been discussed in Chapter 6. Over most of the range of normal vision the visual response is approximately a logarithmic function of the brightness B of the object viewed. Therefore, to produce equal increments of sensation, the steps of brightness ΔB must be related by $\Delta B/B = \text{const.}$ As has previously been pointed out, the relation ceases to be valid at low light levels and over the useful range has a threshold value of about 0.02. The darkest portion of the viewing screen of a Kinescope reproducing a normal picture is not

"black" but has a finite brightness B_0. The available visual contrast range for the Kinescope will, therefore, be

$$\ln \frac{B_m}{B_0},$$

where B_m is the maximum brightness of the screen. The ratio B_m/B_0 is often called the brightness contrast range, as distinguished from the visual contrast range. The contrast range is thus limited by the minimum illumination of the screen, as well as the maximum brightness that can be reproduced. In order to indicate the brightness contrast range that would be desirable for a television screen, Table 12.2 of the contrast range encountered in familiar scenes is given.

TABLE 12.2

Clear sunlight and shadow	100 and up
Normal landscape	30 to 50
Interior with normal artificial illumination	20 to 50
Moving-picture screen	50 to 100
Good photograph	15 to 40

It will be apparent from the wide variation of contrast range listed that the human eye is capable of adapting itself to an extreme variety of conditions. A television picture with a range from 15 to 20 would be acceptable as a good picture, although considerable improvement can be obtained if the range is increased to 100. Increasing the contrast range above 100 results in further improvement, but the improvement is relatively small.

The six requirements listed, referring to size, resolution, brightness, color, flicker, and contrast, are essential and must be met by a viewing tube which is to reproduce a picture of high entertainment value. Besides these essentials, there are other practical considerations. The tube must be such that it can be manufactured economically. Its physical size (as distinguished from screen size) must be reasonable. The voltage and current needed for its operation must be such that they can be cheaply and safely supplied. Finally, the Kinescope should be durable, both electrically and mechanically.

Formidable though this list of requirements may appear, the present Kinescope meets the conditions named in a fairly satisfactory manner.

12.2. Construction of the Kinescope Blank. The design of the blank itself is a major problem in Kinescopes having large viewing surfaces. This is immediately apparent from an estimate of the force exerted by atmospheric pressure on the face of the tube. For example, on the screen alone of a 12-inch Kinescope there is a force of 1700 pounds, or about 7/8

ton. The construction of a glass blank which will stand up under these conditions and yet is not made unduly heavy and costly because of excess material is far from simple. The material of these blanks may be either hard or soft glass. Hard glasses, such as Pyrex and Nonex, have the advantage of excellent thermal and mechanical properties. From the elastic properties of the glass it has been found that, for a 12-inch screen having a radius of curvature of 16 inches, except near the edges of the screen, a wall thickness of ¼ inch is ample; that is, it allows a factor of safety between 3 and 5.

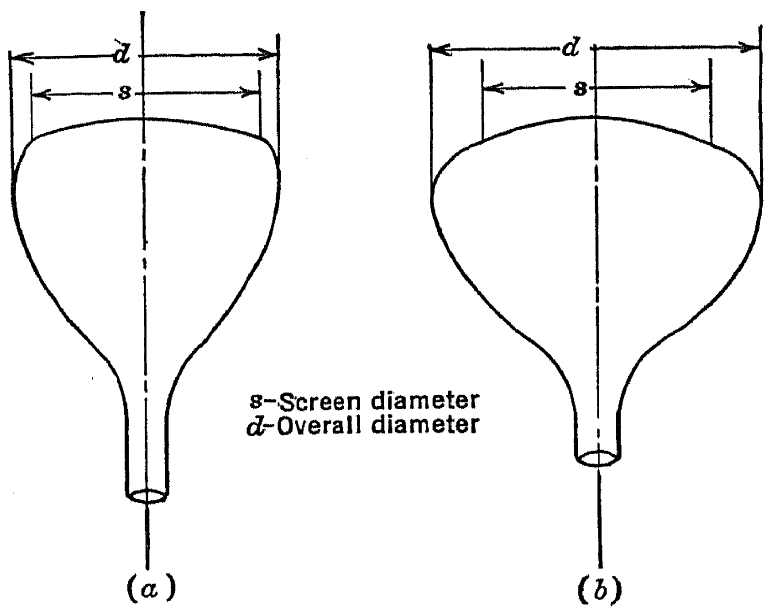


FIG. 12.2.—Two Typical Kinescope Blanks.

At the juncture of the curved disk of the screen and the conical walls of the tube the stresses may become very high. Either the glass must be considerably thickened at this point, or else a graduated curve, which can be calculated by approximate methods, must carry the curvature of the screen into the conical walls. The two types of blanks are shown in Fig. 12.2. The last-mentioned bulb shape requires less glass, but the blank diameter for a given screen size must be somewhat greater. The normal Kinescope combines both principles in order to provide the desired factor of safety.

The conical portion of the tube does not require as great a wall thickness as the screen, 3/16 to 1/8 inch being ample to withstand the stresses. In order to increase the strength for a given amount of material, and for reasons of contrast, the generatrices of the cone are slightly curved rather than straight, as can be seen from Fig. 12.2. The gun neck which is joined to the narrow end of the cone could, as far as the atmospheric pressure is concerned, be made of very thin glass. However, since it is

subjected to other mechanical stresses, and to simplify the sealing of the cone to the neck, it is made the same thickness as the rest of the cone.

Before a blank design is accepted it is put through rigorous pressure tests. The bulb is rough pumped and then placed in a pressure tank. The external pressure is then increased until the bulb implodes. Unless the bulb will withstand from 3 to 5 times atmospheric pressure, the design is rejected.

The blank must be arranged to meet the electrical requirements. In addition to the leads of the gun press, a side contact is provided in the conical walls of the blank through which the high voltage is supplied to the second anode.

The second anode consists of a conductive coating which extends into the gun neck, past the end of the gun. This coating also serves to collect the secondary electrons from the screen and therefore covers the inside walls of the tube, being carried almost to the screen. A metal film like that constituting the second anode of the Iconoscope cannot be used in the Kinescope because the light reflected back on the screen from the brighter portions of the picture would materially decrease the contrast range. For this reason the coating is made of a matte black material. A number of substances are suitable for this purpose, e.g.: Aquadag, Aquagraph, Dixonac, lead sulphide, and carbon black in sodium silicate.

There are many methods of applying the blackening. In experimental tubes, especially those having a settled screen, the following procedure leads to a uniform coating without danger of contaminating the screen with the blackening material. A quantity of the blackening suspension somewhat greater than the volume of the tube to be coated is prepared. If commercial Aquadag is used it should be diluted with about three parts of distilled water. The tube in question is inverted and the gun neck closed with a two-hole stopper. The blackening suspension is admitted through one of the holes, while a glass tube through the second extending to about ½ inch from the screen serves as an air vent. A convenient way of controlling the amount of blackening liquid admitted is to supply the material from a reservoir which can be raised or lowered, and which is connected to the blank being treated by means of a rubber tube attached to an outlet at the bottom of the reservoir. Fig. 12.3*a* shows the arrangement of the equipment. The reservoir is raised until the level of the liquid in the blank just reaches the edge of the screen, then lowered to drain the bulb. Next the coating is dried with a gentle air blast until it is entirely free from moisture. In order to drive off the organic materials used as dispersing agent in the Aquadag, the final step is to bake the tube in air at 400°C. The finished tube will then have a uniform, dull black conducting coating.

This process yields excellent results with experimental tubes, but it is too lengthy and expensive to meet the requirements of large-scale production. Manufacturing methods involve spraying the inside of the blank with special sprayers and masks. With suitable equipment, results equal to, or better than, those with the above-described technique can be obtained. Alternative commercial processes for applying the black coating are the brushing method and the so-called rolling method. For the latter the blank is mounted in such a way that the axis of the tube makes an angle with the horizontal which is slightly greater than the half vertex angle of the conical portion of the bulb. The blackening suspension is flowed into the blank until it just reaches the screen. By rotating the bulb the coating is made to flow uniformly over the inside walls of the tube,

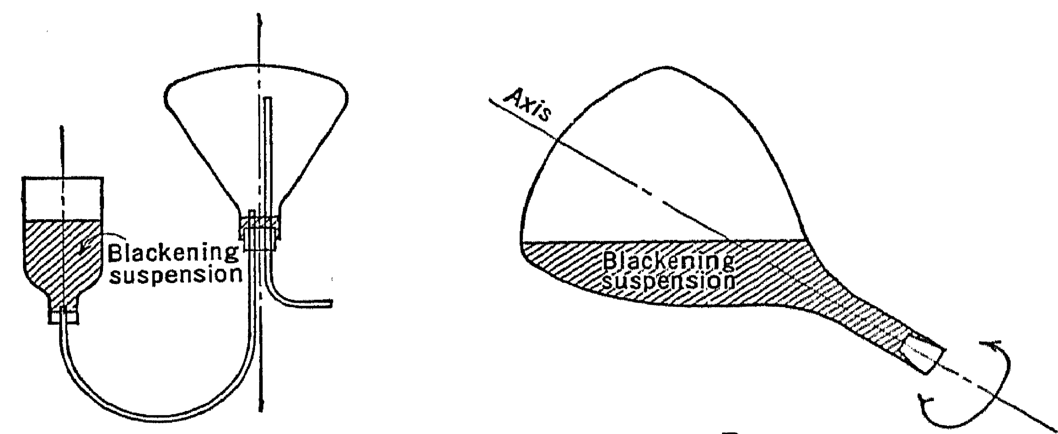


FIG. 12.3.—Methods of Applying Blackening to Form Second Anode.

yet leaving the screen free from contamination. Fig. 12.3*b* illustrates the position of the blank and the coating liquid.

Before leaving the subject of the second anode coating, mention should be made of the method of making contact between it and the lead wire sealed into the wall of the blank. Usually, the lead consists of a short length of wire projecting into the tube. If the black coating is merely deposited over the wire, the contact will be very fragile and will be almost certain to open up during subsequent operation. However, if the wire and the glass for an area of about a half an inch about the lead is coated with silver paste which can be reduced to silver by heating, and the blackening layer is deposited over this silver, a firm, rugged connection between the external lead and the black coating can be effected.

12.3. The Electron Gun. Like the electron gun in the Iconoscope, the Kinescope gun is an electron-optical system for producing a narrow pencil of electrons. The requirements of the Kinescope gun are very different from those of the Iconoscope. It must be capable of producing several hundred microamperes beam current instead of a mere fraction

of a microampere; it must have control characteristics such that it can be driven from the output of the video amplifier; and it usually is operated at a much higher overall voltage. On the other hand, a larger spot size can be tolerated.

The electron power which the gun is capable of delivering is one of the two factors determining the brightness of the reproduced picture, the other being the conversion efficiency of the fluorescent screen material. Because of its importance and because of the rather complicated theory

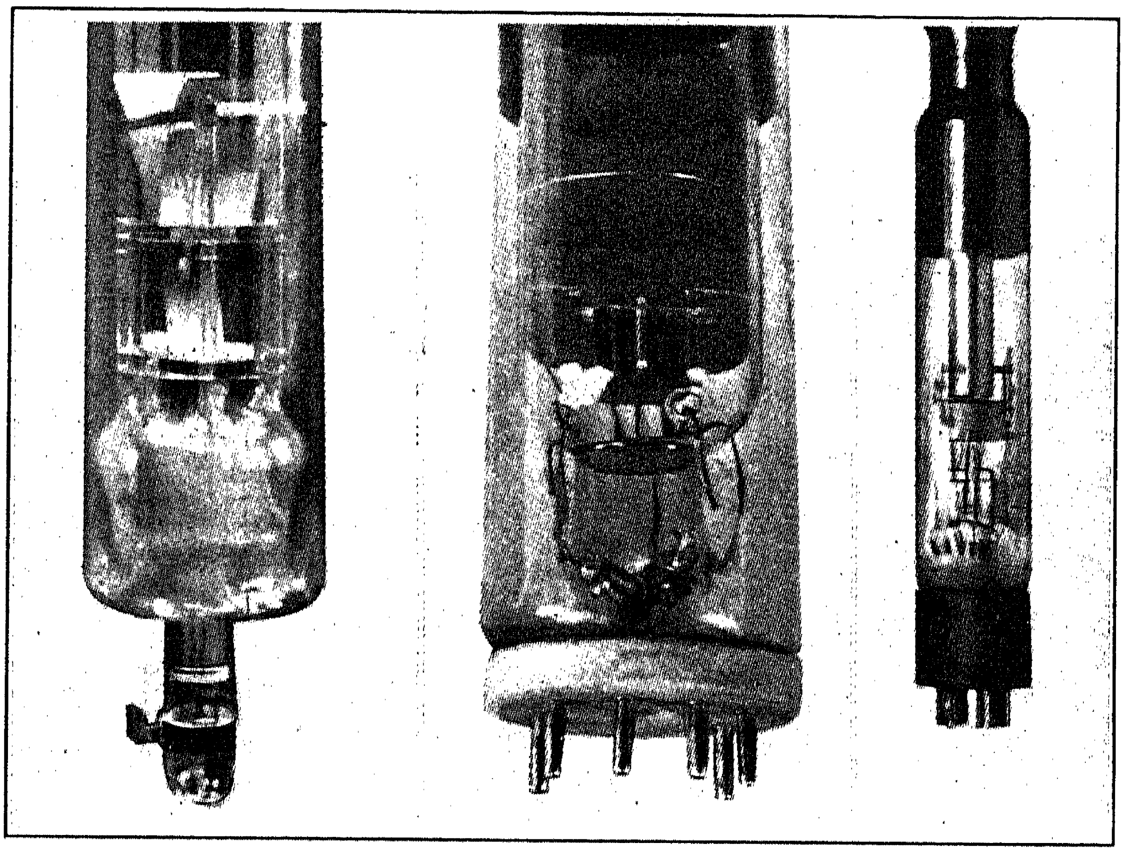


Fig. 12.4.—Three Modern Viewing Tube Gun Assemblies (Courtesy of J. Springer).

back of the modern gun, a detailed discussion of it is reserved for a later chapter dealing with the gun problem alone.

The gun structure is assembled on a press which is sealed into the end of the gun neck in such a way that the axis of the gun coincides with the axis of the tube. At the center of the press a short length of tube serves as the exhaust exit.

Although the processing of the Kinescope does not require the admission of an activating material, nevertheless a getter must be provided to insure long life of the tube. The getter is usually attached to one of the wires of the gun press and mounted close to the press itself. In order to prevent short-circuiting the gun leads, the getter is fastened to a small metal shield which directs the material on the walls of the gun neck. This

shield is also the means of heating the getter when it is exploded with a high-frequency bombarder.

There are, of course, a number of different types of guns, most of them based upon the two-lens system. Three representative electron guns mounted in viewing tubes are shown in Fig. 12.4. These are to be compared with the typical Kinescope gun assembly shown in Fig. 12.5.

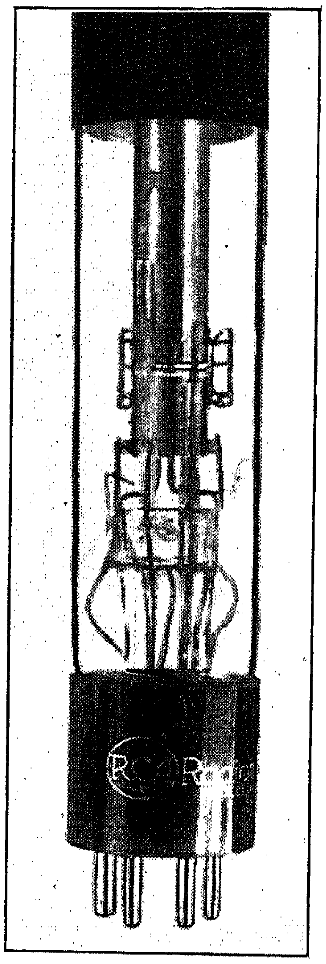


Fig. 12.5.—Typical Kinescope Gun Assembly.

12.4. The Fluorescent Screen. The fundamental importance of the fluorescent screen of the Kinescope is obvious. Its properties determine the color of the image, its brightness for a given beam current, and also, to a certain extent, contrast, flicker, etc. The nature of the fluorescent materials available for coating the viewing screen was discussed in Chapter 2. Of the materials described, those most frequently used are zinc beryllium silicate (having a yellow luminescence) and a mixture of zinc sulphide with either cadmium zinc sulphide or zinc beryllium silicate (producing nearly white luminescence).

The preparation* of the viewing screen does not end with the production of the fluorescent material. The material must be applied to the screen, an operation as important in determining the success or failure of the tube as the actual manufacture of the screen material itself.

As used in the types of Kinescopes discussed in the preceding sections, the screen is viewed from the side opposite that from which the scanning electron beam strikes. The screen consists of a thin layer of finely powdered fluorescent material on the glass end of the tube. The material may be attached to the glass with an inert binder, or it may be ground and deposited in such a way that it adheres without the aid of any foreign substance. The thickness of the layer of fluorescent material used under these conditions is rather critical. If the layer is too thin the electron beam will not be stopped by the fluorescent material but will pass through it, giving up energy to the glass where it serves no useful purpose. On the other hand, a layer which is thicker than the penetration depth of the electrons will absorb some of the light unnecessarily as it passes through the underlying material which is not reached by the beam. Not only the screen thickness but also the size of the particles making up the screen are important in determining its efficiency. From the standpoint

* See Leverenz, reference 8.

of the transmission of light through the screen it can be shown that the size should be as small as possible. However, each individual grain is covered with a "dead layer," i.e., a shell of inert material which does not fluoresce. The thickness of the "dead layer" is approximately independent of the particle size, and the ratio of inert material to active phosphor increases as the particle size is diminished. Therefore, a compromise must be sought in the dimensions of the grains. On the basis of extensive tests and measurements, an average grain size of 0.001 mm has been selected for both zinc beryllium silicate and white sulphide phosphors. The optimum screen thickness can easily be determined once the particle size has been selected. Fig. 12.6 shows the variation of light output with thickness

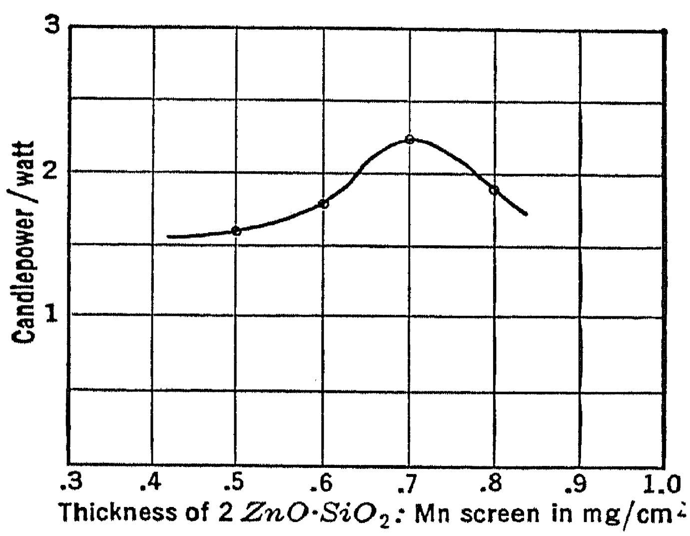


FIG. 12.6.—Light Output as Function of Screen Thickness.

for a settled silicate screen at 10,000 volts. The optimum thickness is seen to be that which results when 0.7 mg of material per square centimeter is used. Since the density of the material is 4.2 gm/cm³, the effective thickness of this screen will be 0.0016 mm. It should be noted that the effective thickness is very different from the actual thickness because the material is not close-packed.

The penetration of electrons through willemite can be shown, by means of relations developed by H. Bethe, to be given by

$$p = 2.5 \times 10^{-12} V^2 \text{cm},$$

where p is the depth of penetration and V the velocity in volts of the electrons striking the screen. While the voltage applied in the measurements shown in Fig. 12.6 was 10,000 volts, the actual velocity of the electrons striking the screen was probably less than 9000 volts, owing to the difference in potential between the second anode and the screen, which was pointed out in Chapter 2. Therefore, it can be concluded that

the optimum thickness is only slightly less than the depth of penetration of the electrons. This is also found to be true for the sulphide phosphors, whose penetrations are given by:

Zinc sulphide.	$p = 2.83 \times 10^{-12} V^2$cm
Cadmium sulphide.	$p = 3.24 \times 10^{-12} V^2$cm

and which require screen thicknesses 10 to 30 per cent greater than willemite for optimum thickness.

As was pointed out in Chapter 2, the break-point potential of a fluorescent screen is, at least in part, dependent on the screen thickness. In the case of the orthosilicate, a screen of 0.7 mg per square centimeter comes under the classification of intermediate thickness screens and has no sharply defined maximum bombarding voltage. Instead the difference between the bombarding voltage and second anode potential increases gradually with rising second anode potential after 4000 to 6000 volts have been exceeded. The continued rise in true bombarding voltage makes it advantageous to operate the projection-type Kinescopes, which will be described in a later section, at very high second anode voltage, that is, 10 to 30 kilovolts or higher.

12.5. Screening Procedure. The size of the freshly synthesized orthosilicate particles may be as large as 0.1 mm when ordinary preparation technique is used. Before the material can be applied to a screen the grain size must be reduced by a factor of about a hundred. Since the efficiency of the material is dependent upon complicated and unstable structural properties, it is very sensitive to mechanical treatment. The grinding necessary to reduce the particle size should, therefore, be done in such a way that the minimum force required to fracture grains comes to bear on the material. This condition can best be met by grinding the phosphor in a ball mill, turned at a rate which produces a sliding rather than a tumbling action. The Moh hardness of the silicate phosphor is 5.5; consequently quartz balls with a hardness of 7 are satisfactory. The material is usually suspended in an inert liquid such as alcohol, acetone, or water. A section through a typical quartz ball mill is shown in Fig. 12.7.

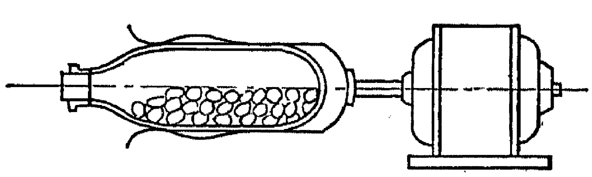


FIG. 12.7.—Ball Mill.

The sulphide phosphors are even more delicate than the silicates, and, in general, it is not practical to reduce their size by grinding. However, by the proper preparation techniques these materials can be crystallized in grains of a suitable size.

The material can be applied to the screen in a variety of ways. Those

most frequently used are: dusting, spraying, and settling through a liquid medium.

The last-named technique permits an accurate control of screen thickness and is, therefore, well suited to experimental production of screens. Furthermore, it wastes no material.

A suspension of the ground fluorescent material in very pure distilled water is made, using an amount of phosphor just sufficient to cover the area of the screen to the desired thickness. The volume of liquid used in making the suspension is relatively unimportant but should be sufficient to cover the screen to a depth of 5 to 10 cm. Since the dielectric constants of the liquid and the particles are different (i.e., for water, $E = 80$; for zinc orthosilicate, $E = 15$), the suspended particles tend to become negatively charged. Such charges will cause the material to settle non-uniformly, producing a screen covered with fine lines which resemble the force lines on a surface dusted with iron filings in the presence of a non-uniform magnetic field. To avoid this difficulty the particles must be kept charge-free, which can be done by adding a small amount of mild electrolyte to the distilled water. Ammonium carbonate has been found to be very satisfactory for this purpose. It should be used in a concentration of about 12.5 grams of the ammonium salt for each gram of the phosphor.

A very convenient procedure is to grind the phosphor in a concentrated aqueous solution of ammonium carbonate. The resulting mixture can be directly diluted to the desired concentration required for the settling suspension. Not only does the slightly alkaline ammonium carbonate solution serve to eliminate the charges on the particles, but also it may be made to reduce the dead layer by a slight decomposition of the material. Further, by a similar action, it increases the adhesion of the phosphor to the screen on which it is settled.

A suitable quantity of the phosphor in suspension having been prepared, the liquid is poured into the blank and the screen material allowed to settle, after which the remaining liquid is decanted and the screen dried. The settling requires from 3 to 8 hours, and during this time the blank must be kept quite motionless and free from vibration. Also, the liquid must be kept at a constant low temperature to avoid undesirable convection currents which will cause the screen to be "loaded" at the center or the edges. If the ambient temperature is lower than that of the liquid, the convection currents will be such that the center of the screen will be too thick, while a higher ambient temperature produces the opposite effect. It should be noted that small, controlled convection currents can be used to obtain a uniform screen on a curved surface such as is usually encountered in a normal Kinescope.

When the settling is complete, no scattering should be observable from a light beam passed through the liquid parallel to the screen. The liquid can then be decanted slowly over a period of half or three-quarters of an hour. During this pouring process, vibration must be avoided. After the pouring has removed all the liquid, the screen is dried with a gentle current of clean warm air.

In order to carry out the settling and pouring operation described it has been found convenient to use the apparatus illustrated in Fig. 12.8. The blank is clamped to a small platform which is hinged to a massive cast-iron base. The base is suspended by means of damped helical springs, so that it will be free from vibration. A smooth-running motor is arranged so that it will slowly tilt the hinged platform and the Kinescope clamped to it, in this way performing the pouring operation.

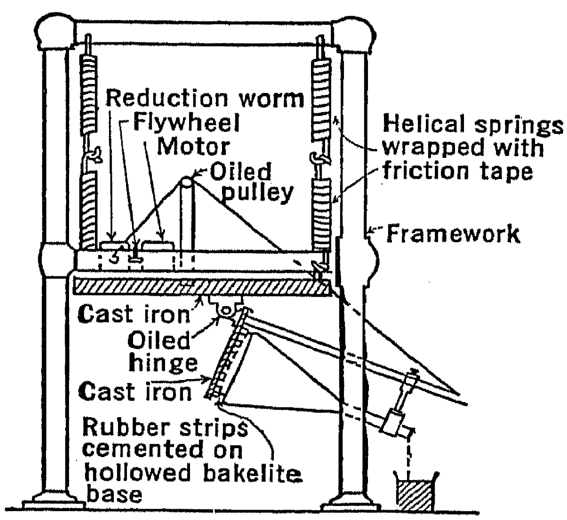


FIG. 12.8.—Settling and Pouring Apparatus.

The settling operation just described is to some extent applied to the production of normal direct-viewing tubes, but is in general considered rather lengthy. Projection Kinescope screens are frequently made by this process.

Application of the screen material by spraying lends itself to production methods. In this case the phosphor is often suspended in a volatile organic liquid, such as acetone, to which has been added a small amount of binder. The binder may be in the form of nitrocellulose or cellulose acetate, or some similar material which can easily be removed by baking after the screen has been applied. In general, however, a binder is not necessary, as the screen material will adhere if the particle size is small. To avoid the charging of the phosphor grains it is often necessary to add a small amount of organic acid (e.g., formic acid or acetic acid) or base (e.g., trimethylamine) to the suspending liquid. The particle size suitable for spraying is from 1 to 6 microns.

Larger particles, from 10 to 30 microns in diameter, are usually deposited on the screen by dusting or settling through air. A binder in the form of sodium or potassium silicate is often required to cause the material to cling to the screen. This technique lends itself well to production methods and is extensively used by tube manufacturers. The powdered material is put into a dust gun which is inserted in the inverted Kinescope. The gun produces a uniform cloud of the phosphor which settles

on the glass end of the tube. Two forms of the dust gun are shown in Fig. 12.9. Because it lends itself to the deposition of large particles, dusting is widely applied in the production of sulphide phosphor screens, which are easily damaged by grinding.

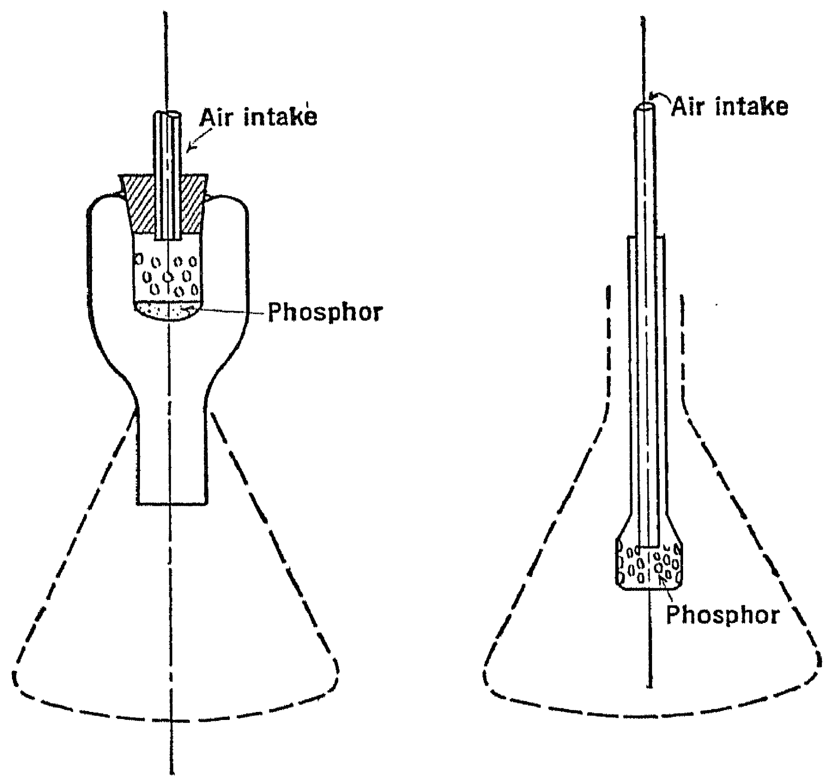


FIG. 12.9.—Dust Guns for Applying Screen Material.

12.6. Processing of the Kinescope. The experimental exhaust and activation of a Kinescope is relatively simple. An ordinary vacuum system capable of producing a pressure of 10^{-6} mm Hg or less, and equipped with an oven, is satisfactory. The evacuation schedule is the same as that for the Iconoscope, the tube being baked on the pump at a temperature of 350°C to 450°C, depending upon the kind of glass used in the blank. After the tube has been baked the gun parts are thoroughly outgassed by heating them to red heat for a period of 10 minutes to half an hour. The activation of the gun, which then follows, proceeds exactly as described in Chapter 1. Just before the tube is sealed off the system the getter is exploded by heating the metal tab to which it is attached with the high-frequency bombarder.

A tube prepared in this way should give good service and have a long life. The principal precaution to be taken is to avoid residual gas. The beam current in a Kinescope is relatively high, and therefore even a small amount of gas will result in the generation of positive ions which will greatly decrease the life of the thermionic cathode. For this reason

the walls of the tube and the metal gun parts must be very carefully outgassed.

Residual gas manifests itself first by producing a dark, inactive spot in the center of the fluorescent screen. This spot is caused by the bombardment of the screen with negative ions which destroy the phosphor. While the presence of negative ions does relatively little harm (the dark or discolored area being very unobtrusive and generally not noticeable unless looked for), negative ions are sometimes accompanied by positive ions which may be destructive to the life of the cathode. However, the screens, particularly when sulphide phosphors are used, are very sensitive to ion bombardment, and the appearance of a dark spot even in a new tube has no significance as far as the life of the tube is concerned. Near the end of the life of a tube which has shown good service the dark spot may become quite pronounced.

The nature of the gases producing these effects has been studied by Knoll and others by a rather ingenious method. Before the screen material is destroyed, the ions are capable of producing fluorescence. If the tube is placed in a strong magnetic field with a known flux density, the electron beam can be deflected off the screen while the gas beam will be deflected by an amount depending upon the e/M ratio. Thus the tube itself serves as a mass spectrograph. By reversing the potentials on the tube, positive ions will produce a similar gas beam. The gas beam will usually be broken up and produce several spots showing the presence of more than one gas. By measurement of the positions of these spots the gases have been identified. The principal gases found in order of their occurrence by volume are: oxygen, hydrogen, carbon monoxide, water vapor, nitrogen, and carbon dioxide. Other gases are present but in lesser quantity.

The production processing of the Kinescope presents a rather different problem. Here time and economy of operations are of major importance. Particularly to be avoided are all hand-performed individual operations which require skilled workers. Therefore, the development of the technique of processing the Kinescope on Sealex automatic machinery has constituted a major advance and has greatly reduced the cost of production. The machine resembles the pumping unit of the ordinary Sealex machine used in the commercial handling of vacuum tubes. The tubes to be processed are mounted vertically around the edge of a large wheel which turns in a horizontal plane. The exhaust stem of each tube fits into a soft-rubber socket and is clamped so that the joint is vacuum tight. Each socket is directly connected to a mercury vacuum pump backed by a common fore-pump. The wheel is rotated in steps, each step moving the tubes into position for a new operation. In all there are six-

teen positions carrying out five different operations. These operations are:

1. Initial exhaust.
2. Exhaust and bake.
3. Outgassing of metal parts.
4. Activation of gun.
5. Gettering and seal-off.

The baking time is reduced to an absolute minimum and, while the exhaust is accomplished with the aid of mercury diffusion pumps, no freezing-out traps are used. The success of this procedure rests on the selection of suitable materials for the components of the tube and the copious use of a very active getter. Tubes correctly processed on the automatic machine are as efficient and as long-lived as tubes which have been handled by the long and rather delicate experimental method first described.

After the tube is sealed it is based and the second anode lead cap fastened in place. The tube is then put on an aging rack and operated for a short period.

12.7. Tests and Performance. Every responsible manufacturer must test the viewing tubes carefully to keep a constant check of the production methods. These tests can be divided into two classes. The first class consists of routine checks on the performance of every tube produced. The second consists of a more critical examination of a representative selection of tubes, made periodically.

The routine tests must determine whether the screen on each tube is satisfactory, whether the spot meets the requirements of size and shape, and whether the control characteristic is correct. The tests also should be of such a nature that they can be performed simply and quickly. One form of test outfit which meets these requirements has been designed by C. E. Burnett.* This equipment, in addition to supplying the anode voltages and vertical and horizontal deflection to the tube being tested, is arranged to deliver a signal to the control grid which breaks the pattern up into dots of about picture-element size which vary in brightness over the screen. A pattern of this type is shown in Fig. 12.10. By comparing this pattern with a standard pattern, the performance of the tube can readily be judged. Visual inspection will reveal whether or not the spot dimensions will give the desired resolution. A comparison of the brightness of the pattern with the standard indicates the merit of the screen. The control characteristics can be estimated from the variations of pattern brightness over the screen. With this type of equipment the

* See Burnett, reference 9.

tests can thus be performed simply by inserting the Kinescope into its socket and examining the pattern.

The second class of examination is much more rigorous and must supply quantitative information about the performance of the Kinescope not only under normal operating conditions but also under conditions of overload.

Tests of the electron-gun performance should include a microscopic examination of the spot in various positions on the screen, and microm-

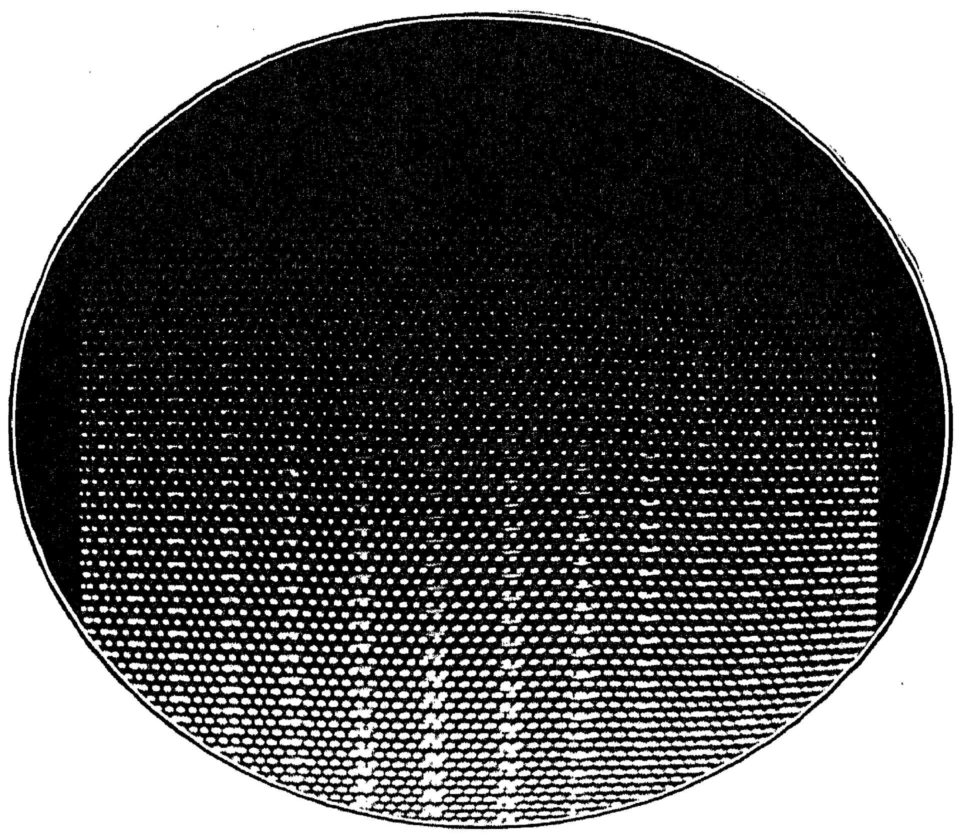


FIG. 12.10.—Kinescope Test Pattern.

eter measurements of spot size. The line produced by deflecting the beam is similarly examined. The extent to which the spot "blooms" or increases in size when the grid is made positive is also determined.

Furthermore, the cathode should be studied for flaws in the emitting coating. This can be done by adjusting the potentials on the gun in such a way as to produce a magnified image of the cathode on the fluorescent screen. In other words, the gun is used as an electron microscope.

In addition, measurements are made of the control characteristics of the gun and of the current to each of the gun elements. Overvoltage tests of the gun yield information as to the quality of the insulation and the extent of cold discharge.

The fluorescent screen should be examined to determine quantitatively

the efficiency of the phosphor, and curves taken of light output against current and voltage. Also the spectral output should be measured.

Finally, the tubes are life-tested, that is, run under normal operating condition until they fail, in order to determine the operating life of the tubes and the nature of failure.

12.8. Contrast. The behavior of the Kinescope with respect to contrast is an extremely important factor in determining its overall performance. Therefore, the causes which contribute to loss of contrast must be studied in some detail. The problem is not a simple one, and, though recent studies* have made it possible to construct tubes with a reasonably high contrast ratio, a complete solution has yet to be found.

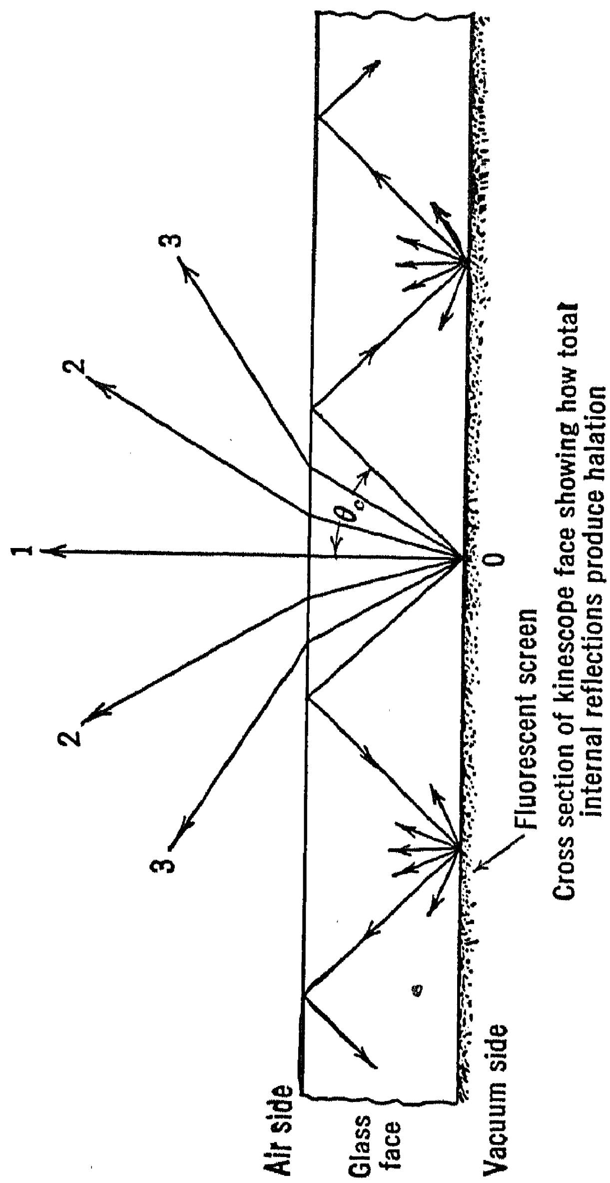
If a conventional Kinescope is observed when the spot is stationary at the center of the viewing screen it will be noticed that, in addition to the single bright point of light due to the impact of the electron beam, considerable light is scattered from other parts of the screen. Close to the beam, and extending over an area of several square centimeters, is a rather faint luminosity due to normal reflections. The spot is surrounded in addition by two or more concentric rings of light, the smallest having a radius of about a centimeter, due to total internal reflections in the glass face of the tube. Finally, there is a diffused glow over the entire screen which is more or less uniform but may have a slight maximum in the vicinity of the spot. This general scattering is due primarily to the effect of screen curvature, reflections from the bulb walls, and scattered electrons.

A consideration of these factors will make it immediately apparent that, for a given tube, the value of contrast range which can be obtained depends upon the type of pattern reproduced. In other words, much greater contrast is possible when the pattern consists of a small bright area on a dark background than when it consists of a dark area on a white background. Therefore, a single figure will not suffice to describe the performance of the tube with respect to contrast. Two measurements which, when taken together, give a fairly accurate estimate of the merit of the tube, are those giving the detail contrast range and the overall contrast range.

The detail contrast range is measured by observing the brightness ratio of a small dark area on a bright background. For this measurement the beam is made to scan the normal picture area and a signal is applied to the control grid such that the beam has its maximum operating intensity except over an area which is small compared to the picture area, where the beam is biased to cutoff.

If the same area is scanned, but a signal is applied to the grid which

* See Law, reference 10.



Cross section of kinescope face showing how total internal reflections produce halation

(*a*)

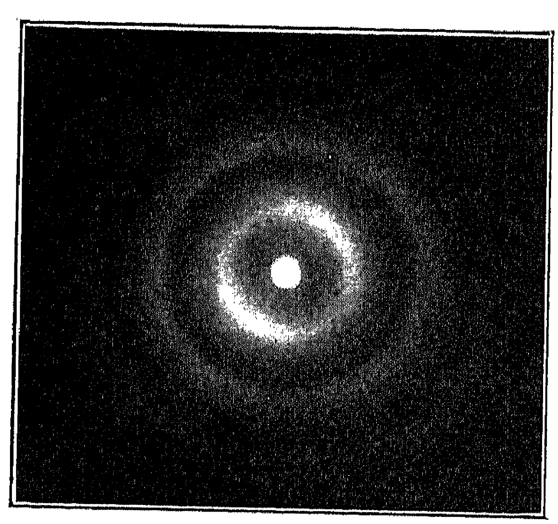


FIG. 12.11.—Internal Total Reflection.

biases the beam to cutoff over half the picture, the overall contrast range can be obtained by measuring the brightness at the centers of the light and dark areas.

The major factor in reducing the contrast range of a picture is the scattering due to internal reflections in the glass. Fig. 12.11 illustrates the nature of this type of scattering. The screen material is shown as being embedded in the glass wall. Rays leaving the screen at the point of electron bombardment will be refracted at the interface between the glass and air. Any ray which leaves the screen material at an angle to the normal greater than θ_c will not emerge from the glass but instead will be totally reflected back to the screen. When this totally reflected light again strikes the screen material some of it will be scattered, producing the first visible ring, while the remainder will be totally reflected to form additional rings.

The critical angle required to produce total reflection can be calculated, from Snell's law and a knowledge of the index of refraction μ_1 of the glass, as follows:

$$\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2,$$
$$\mu_2 = 1, \ \mu_1 = 1.54,$$
$$\theta_2 = 90°, \ \theta_1 = \theta_c,$$

therefore:

$$\theta_c = \sin^{-1} \frac{1}{1.54} = 41°.$$

From the geometry of Fig. 12.11a it will be seen that the diameter of the first ring will be given by the relation:

$$D = 4 \cdot t \cdot \tan \theta_c,$$

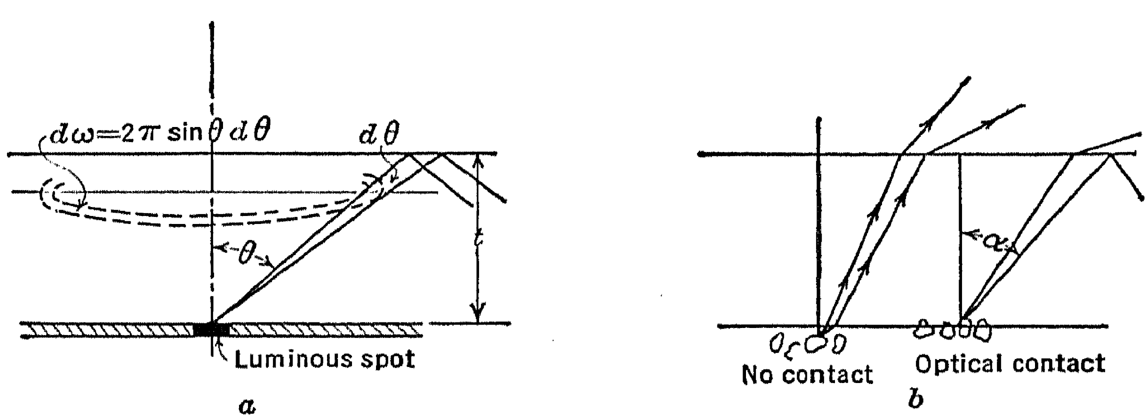


Fig. 12.12.—Reflection and Refraction in the Tube Face.

where t is the thickness of the glass and can be assumed to be ¼ inch. This gives a diameter of 0.85 inch.

It is evident that only a fraction of the light emitted at the spot

escapes from the screen in the form of useful light. Since the distribution of light emitted by the fluorescent screen approximates Lambert's law:

$$L = B \cos \theta,$$

where B is the light flux per unit area emitted by the screen and θ the angle between the direction of observation and the normal, the fraction of light which can escape can be calculated with the aid of the following integrals, whose construction can be seen from Fig. 12.12a.

$$R = \frac{2\pi \int_0^{\theta_c} B \cos \theta \sin \theta d\theta}{2\pi \int_0^{\pi/2} B \cos \theta \sin \theta d\theta}$$
$$= \sin^2 \theta_c = 42\%.$$

Thus it will be seen that under the condition postulated about 60 per cent of the emitted light is trapped by internal reflection.

The assumption was made, however, that all the phosphor grains were embedded in the glass. Actually, the optical contact between the fluorescent material and the glass is only about 20 to 30 per cent, depending upon the way in which the screen is prepared. As can be seen from Fig. 12.12b, no internal total reflections exist in the case of luminescent grains which are not in optical contact with the glass. Therefore, assuming 30 per cent contact, the actual loss by internal reflection is 60 per cent $\times$ 30 per cent, or 18 per cent of the emitted light. The internally reflected light is gradually scattered out of the screen as it is reflected back and forth between inner and outer glass surfaces. At least half of the scattered light must be in the forward direction, that is, 9 per cent, and contributes to loss of contrast. On this basis, and owing to this cause alone, a contrast ratio not greater than 82 per cent/9 per cent = 9 can be expected. A more accurate calculation, including multiple scattering within the screen, absorption of the glass, and edge effect, reduces this figure to about 6.

Since most of the loss of contrast is due to the scattering at the first ring of light, it has a much greater effect on detail contrast ratio than on overall contrast ratio, a fact which is confirmed by measurements.

Light which is scattered by the curvature of the screen is next in importance in producing loss of contrast. The explanation of this type of scattering is simple. The emitting point of the phosphor is on the inner surface of a spherical shell, and thus every point of the screen can "see" the emitting area. Sidewise emitted light strikes the screen directly and

is scattered in all directions, as shown in Fig. 12.13. The light spread over the screen can be readily calculated. If D is the diameter of the screen and R its radius of curvature, then the maximum angle over which light can reach the screen measured from a tangent to the shell at the

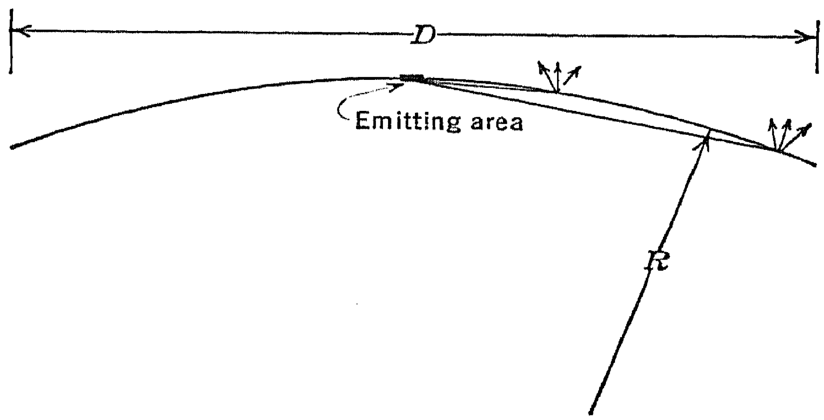


FIG. 12.13.—Effect of Face Curvature on Contrast.

point of impact of the beam (assumed to be the center of the screen) will be α, where:

$$\tan \alpha = \frac{D}{4R} = \sin \alpha.$$

Proceeding as before, the ratio of scattered light to forward light will be:

$$S = 1 - \int_0^{\frac{\pi}{2}-\alpha} 2 \sin \theta \cos \theta d\theta$$
$$= \sin^2 \alpha = \frac{1}{16}\left(\frac{D}{R}\right)^2.$$

Since at least half of the light scattered by the screen is in the backwards direction, the ratio of brightness of the dark area to the bright area (assuming Lambert's law) is no greater than:

$$S = \frac{1}{32}\left(\frac{D}{R}\right)^2.$$

Assuming the values $D = 12$ inches, $R = 16$ inches, S is found to be approximately 0.02, indicating a contrast ratio of 50. This figure is in fairly close agreement with the measured value for this type of scattering.

In addition to the total reflection occurring in the glass face of the screen, normal reflection takes place for angles less than the critical angle. The resulting scattered light cannot readily be calculated with any degree of accuracy. Estimates have been made, however, which show this effect to give a contrast ratio of 60.

Loss of contrast due to reflections in the bulb is small compared with that due to the above-described processes. Bulb reflections may be either specular or diffuse. If the blank is so shaped that specular reflection from one part of the screen reaches other screen areas, the loss of contrast can be quite great. The bulb shown in Fig. 12.14*a* gives this type of reflection.

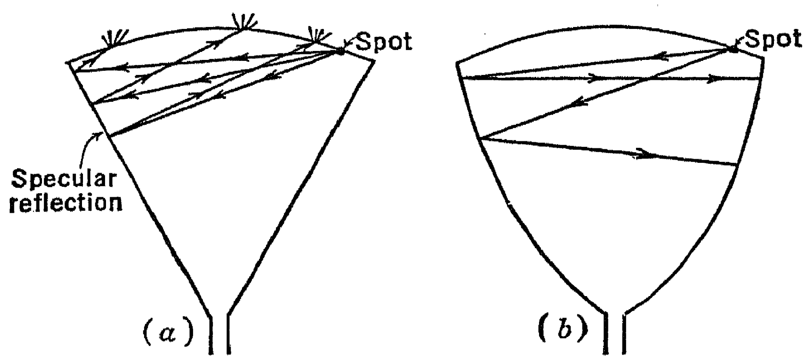


Fig. 12.14.—Effect of Wall Shape on Bulb Reflections.

If the conical part of the blank is curved as shown in Fig. 12.14*b*, specular reflection cannot reach the screen. In the blank in Fig. 12.14*a*, which does not have this curvature, the texture of the blackening material used on the walls is very important and even a surface which is only moderately glossy will greatly reduce contrast. Where the walls are curved to prevent specular reflection, the type of coating is not critical. With such a blank, good contrast can be obtained even if the glass is left uncoated, though this is disadvantageous from an electrical point of view. It should be pointed out that smaller tubes, constructed so that the conical portion has a narrow vertex angle, will have a relatively small area over which specular reflection can take place, as can be seen from Fig. 12.15. For this reason they are generally made with straight walls.

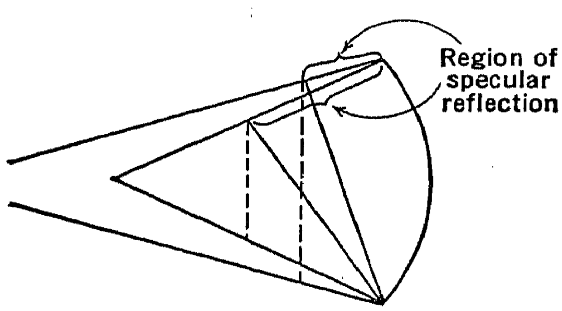


Fig. 12.15.—Bulb Reflection and Cone Angle.

Two sources of scattered electrons may cause a loss of contrast. The first is the electron gun itself. Electrons may come from the gun in such a way that they do not form part of the scanning beam owing either to secondary emission from the gun element or to cold discharge. With the modern gun equipped with apertures to suppress secondary emission, and all edges about which a high gradient might appear, rounded, this source of scattering is negligible as far as loss of contrast is concerned. Electrons scattered by the screen itself may return to the screen because of the potential gradient which exists over its surface. The nature of this

gradient was discussed in Chapter 2. At voltages ordinarily used with the direct-viewing Kinescope, the reduction in contrast due to this second type of scattering is also small.

As a result of the combined losses the contrast ratio observed in a normal Kinescope is:

Detail brightness contrast ratio.	6 – 10
Overall brightness contrast ratio.	30 – 50

12.9. Regular Kinescopes. For the ordinary console television receiver, viewing tubes with a 12-inch screen have been found very satisfactory.

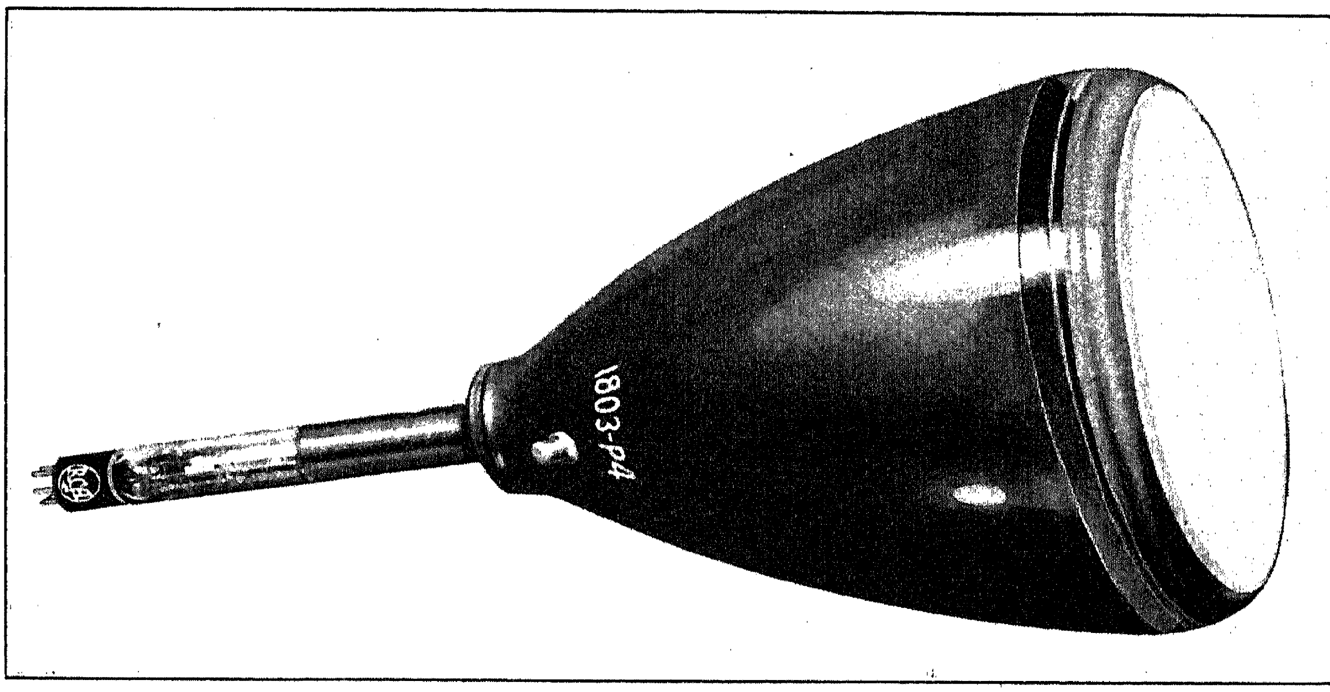


Fig. 12.16.—Twelve-Inch Kinescope RCA 1803-P4.

A typical tube of this size is shown in Fig. 12.16. The grid characteristic in terms of light output is given in Fig. 12.17. Operating specifications for this tube are listed below:

1803-P4 Kinescope

Heater voltage.	2.5	2.5	volts
Anode No. 2 voltage.	6000	7000	volts
Anode No. 1 voltage (approx.)*.	1240	1460	volts
Grid No. 2 voltage.	250	250	volts
Grid No. 1 voltage†‡.	Adjusted to give suitable luminous spot		
Grid No. 1 signal voltage§.	25	25	volts

* Adjustable to ±20%.
† Approximately 30% of grid No. 2 voltage is required for current cutoff.
‡ Maximum resistance in the grid circuit should be limited to 5 megohms.
§ Peak-to-peak value for optimum contrast.

Fig. 12.18 shows a similar type of viewing tube having a 12-inch screen made by Electric and Musical Industries, Ltd., of England.

Smaller tubes are also available, designed especially for television work. Such tubes are used in smaller commercial receivers and are also marketed to meet the demand of experimenters who wish to build their own sets.

Representatives of this type of tube are shown in Figs. 12.19 and 12.20. The RCA 1804-P4 has a 9-inch screen diameter and an overall

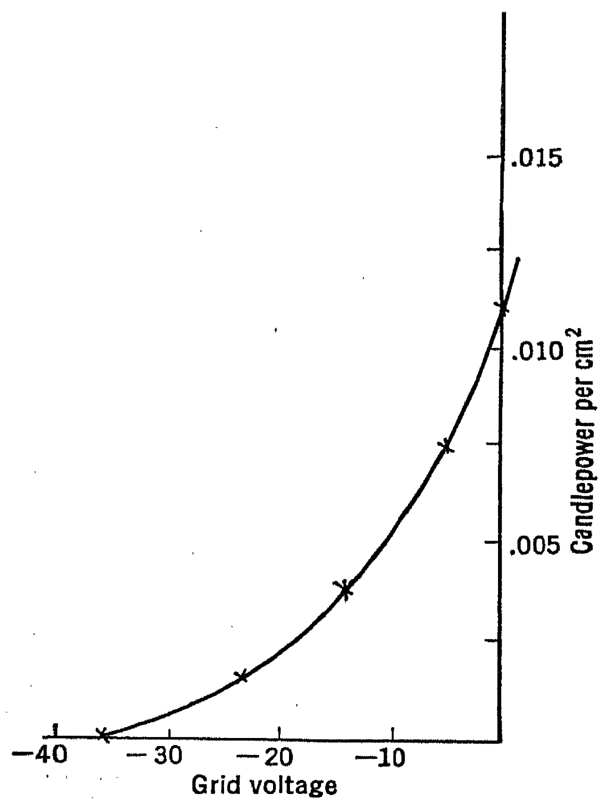


FIG. 12.17.—Grid Characteristic in Terms of Light Output.

length of 21 inches. It is intended for the reproduction of a picture 7⅜ by 5½ inches in size. The gun is electrostatically focused, and both vertical and horizontal deflections are magnetic. Typical operating conditions for the RCA 1804 are:

	Volts	
Heater voltage	2.5	2.5
Second-anode voltage	6000	7000
First-anode voltage	1250	1425
Accelerating-electrode voltage	250	250
Control-grid swing	25	25

The control characteristics are shown in Fig. 12.21. Tube RCA 1802 is much smaller and operates at lower overall voltages. Its diameter is

5 inches and overall length 15¾ inches. In this tube both focusing and deflection are electrostatic. Both tubes are available with a white phos-

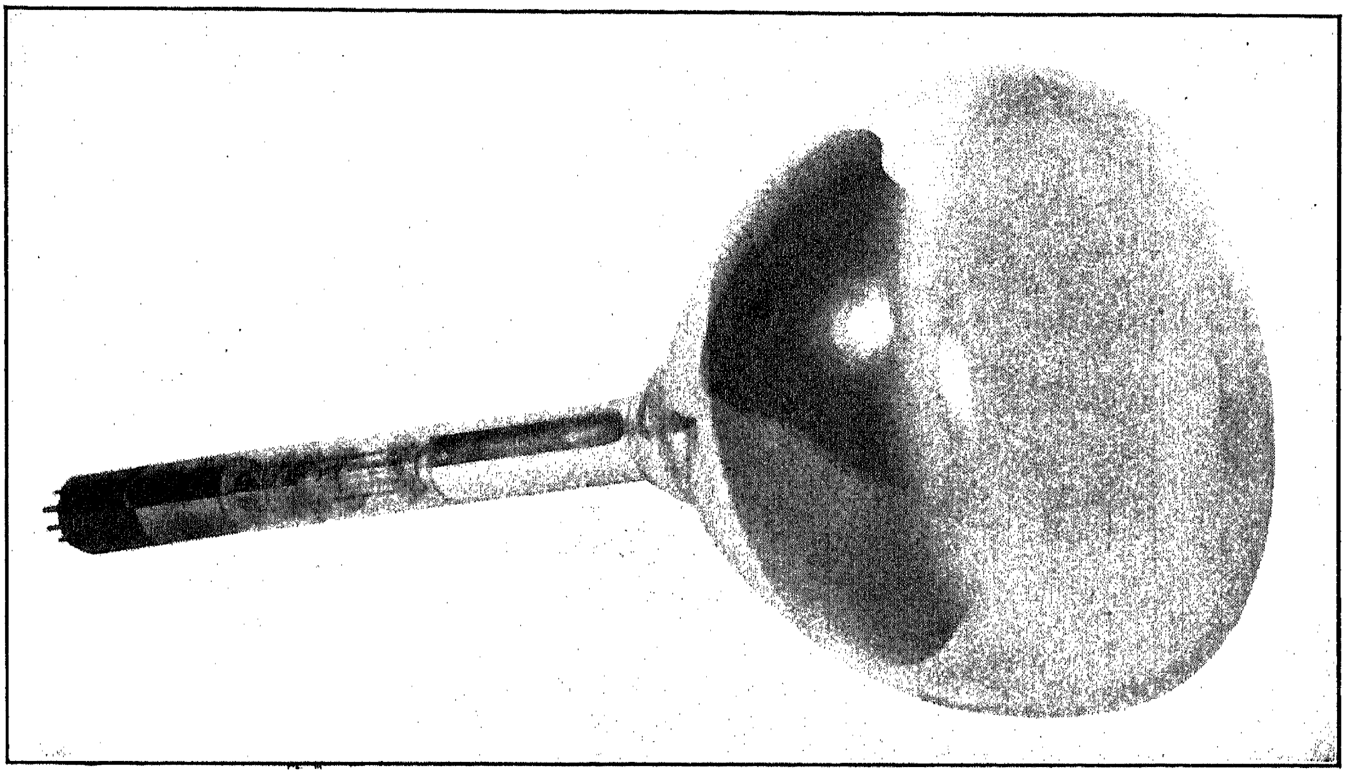


FIG. 12.18.—Twelve-Inch Viewing Tube (Electric and Musical Industries, Ltd., London).

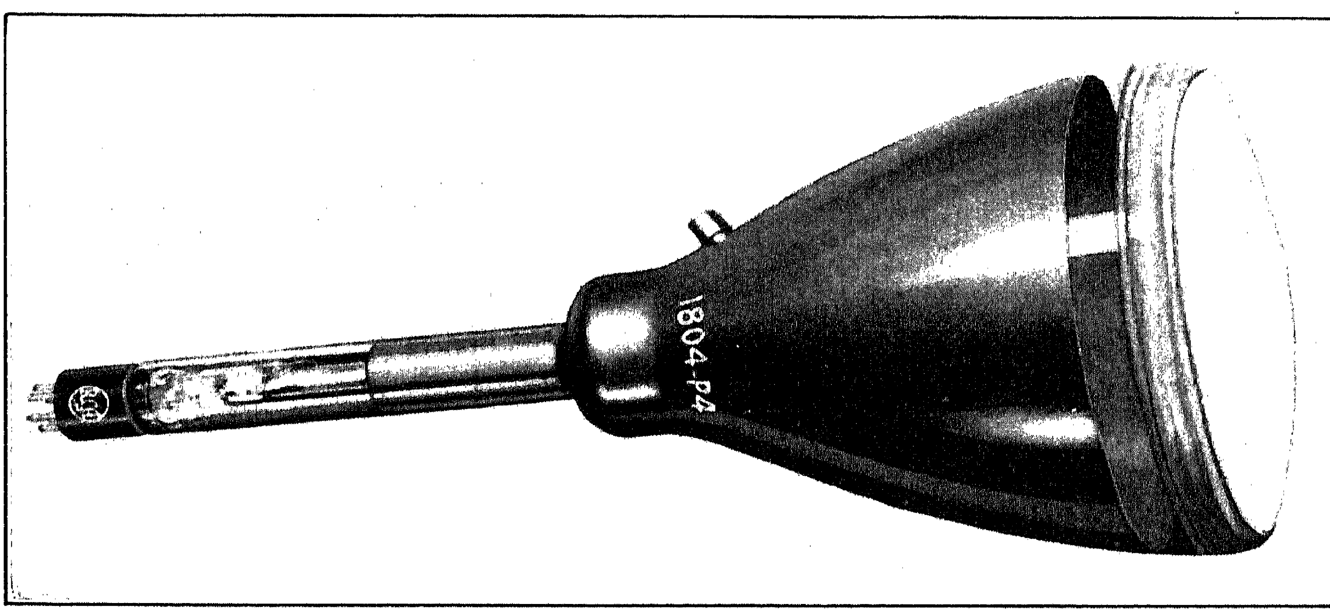


FIG. 12.19.—Nine-Inch Kinescope (RCA 1804-P4).

phor. The screen characteristics are shown in Fig. 12.22; Fig. 12.23 gives the spectral output of the phosphor.

The voltage supply for the RCA 1804 may be a simple, half-wave

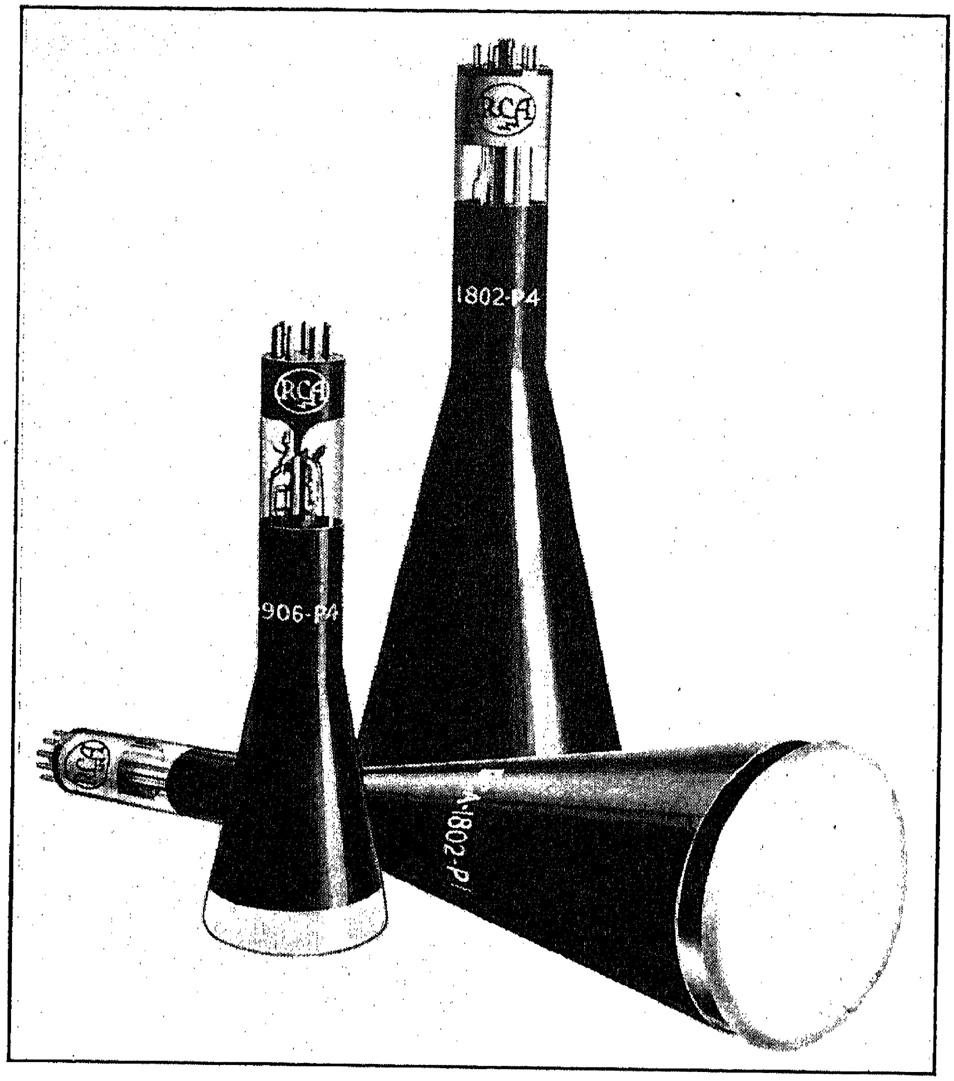


FIG. 12.20.—Small Viewing Tubes Suitable for Television Purposes (RCA 1802-P4, RCA 1802-P1, RCA 906-P4).

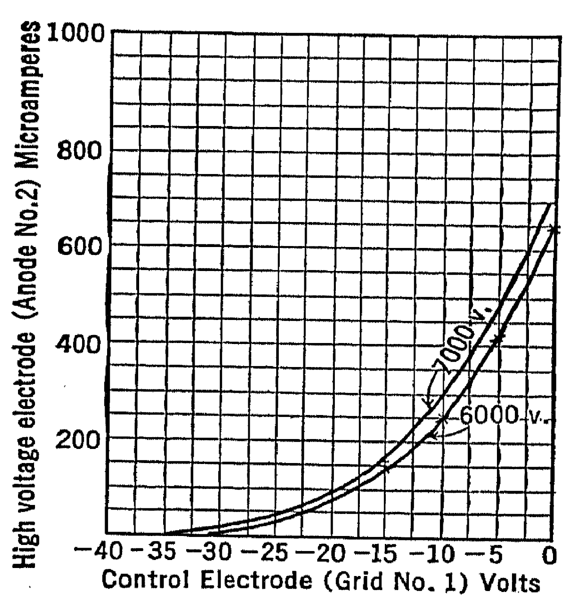


FIG. 12.21.—Control Characteristics of the RCA 1804-P4.

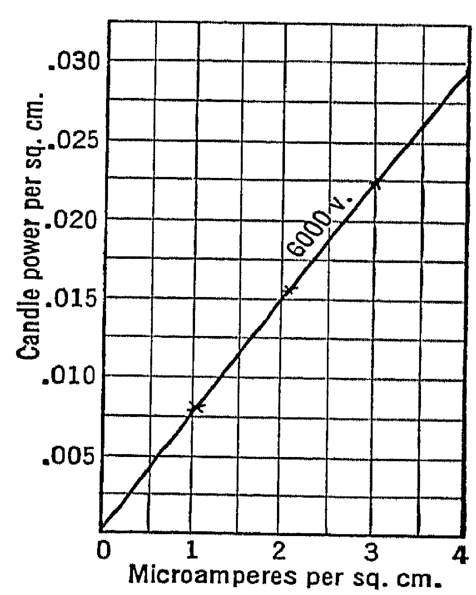


FIG. 12.22.—Light Output as Function of Current Density for White Phosphor No. 4.

rectifier feeding a voltage divider. A circuit of this type is shown in Fig. 12.24.

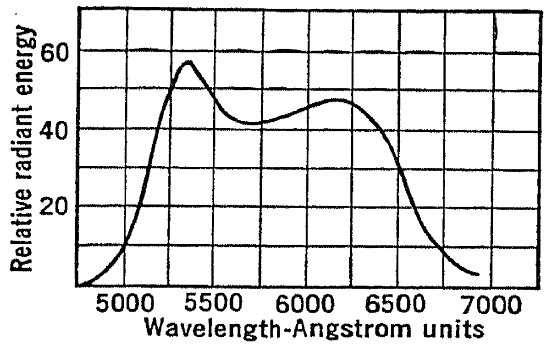


FIG. 12.23.—Spectral Characteristics of White Phosphor No. 4.

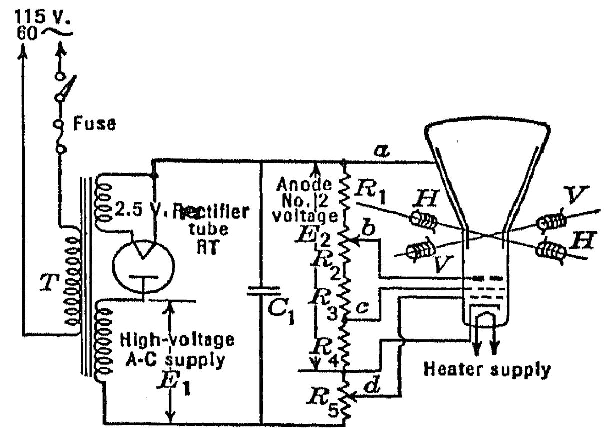


FIG. 12.24.—Simple Voltage Supply for the RCA 1804-P4.

12.10. Special Kinescopes. The improvement of the Kinescope is one of the most important fields of research in the realm of television. There are many lines along which this research may be carried out. In particular, the brightness of the picture, ease of control, contrast ratio, and picture size are subjects of investigation. The first two items involve chiefly improvements of the fluorescent material and gun design, and are discussed in other chapters. Contrast ratio has been very greatly increased in special experimental tubes with screens utilizing principles found in the study of scattering processes. Some of these tubes and screens will be discussed as illustrative of the approach to the general problem.

The detail contrast ratio, as has already been pointed out, is largely limited by internal total reflection. By mixing a small amount of black absorbing material with the phosphor or binder the ratio can be greatly improved. Although such material slightly reduces the efficiency of the screen, nevertheless, since the light producing halation circles must travel much farther through the material than the direct light, there is a large net gain in contrast. In experimental tubes* having a quantity of absorber added to the binder sufficient to reduce the transmission of light through the screen by 10 per cent, the detail contrast ratio is found to be approximately doubled. When 20 per cent of absorbing material is used even greater contrast ratios are attained.

Another type of experimental tube which gives an even greater contrast range is known as a front-viewing Kinescope. This tube differs from the ordinary Kinescope in that the screen is viewed from the same side

* See Law, reference 10.

that is bombarded by the scanning beam. The phosphor is deposited on a flat, opaque surface such as a metal plate or a metallized glass or mica sheet. This screen is mounted in the tube in such a way that the axis of the gun makes an angle of 30° to 45° to the normal of the surface, thus leaving the viewing screen unobstructed. This, of course, requires keystoning the deflection to correct for the angle of the gun. Overall contrast ranges of well over a hundred have been obtained with this type of tube, the contrast being limited by bulb reflections and electron scattering. A practical objection to this type of tube is its awkward geometrical

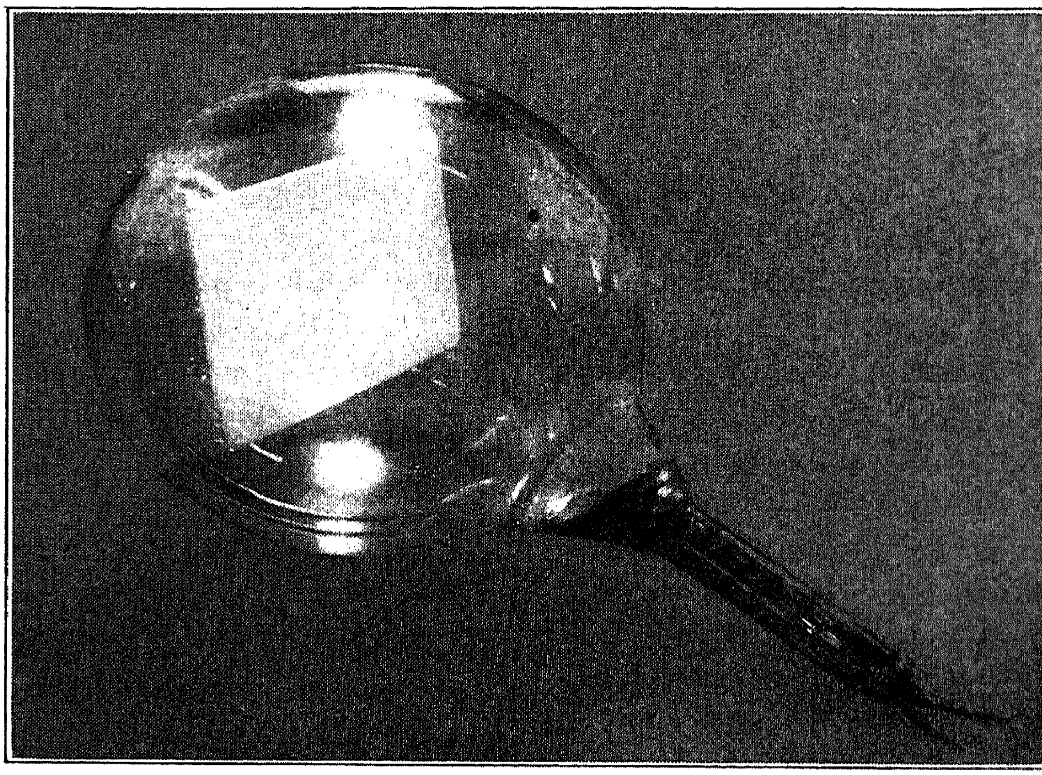


FIG. 12.25.—Front-Surface Kinescope.

arrangement, which makes it difficult to adapt such a tube to a console receiver. A front-vision tube is shown in Fig. 12.25.

A high-contrast tube can also be made, without the difficulty of the physical arrangement of the preceding tube, by incorporating another special screen configuration. This tube makes use of a very thin, flat rear-vision screen separate from the walls of the blank. By making the supporting screen only a few thousandths of an inch thick, as is possible with mica, the size of the total reflection rings can be made sufficiently small so as not to impair contrast. A diagram of this type of Kinescope is shown in Fig. 12.26.

The 12-inch screen of the tube in ordinary receivers is only just large enough to be acceptable for home use. A picture to be viewed by a sizable audience must be larger. The size can be increased by projecting the pic-

ture, as will be explained in the next section. An alternative is offered in the large demountable Kinescope designed by I. G. Maloff.* This tube has a 2-foot screen and an overall length of about 4½ feet. The body of

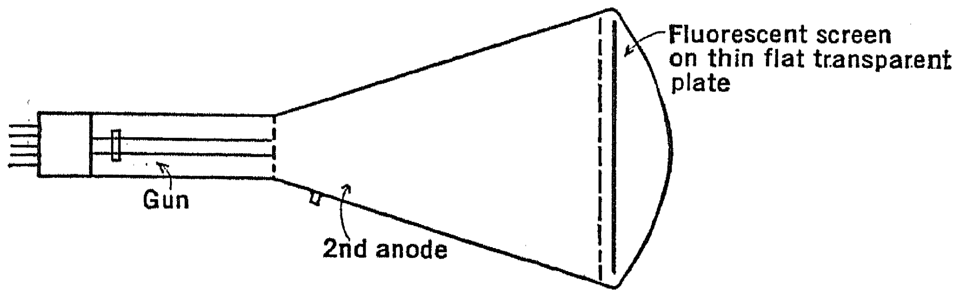


Fig. 12.26.—Kinescope with Separate Thin Flat Viewing Screen.

the tube is a large metal cone which is flanged at both ends. The electron-gun assembly clamps to the smaller flange and is sealed with a gum rubber gasket. Although a peak current of 8 milliamperes can be obtained

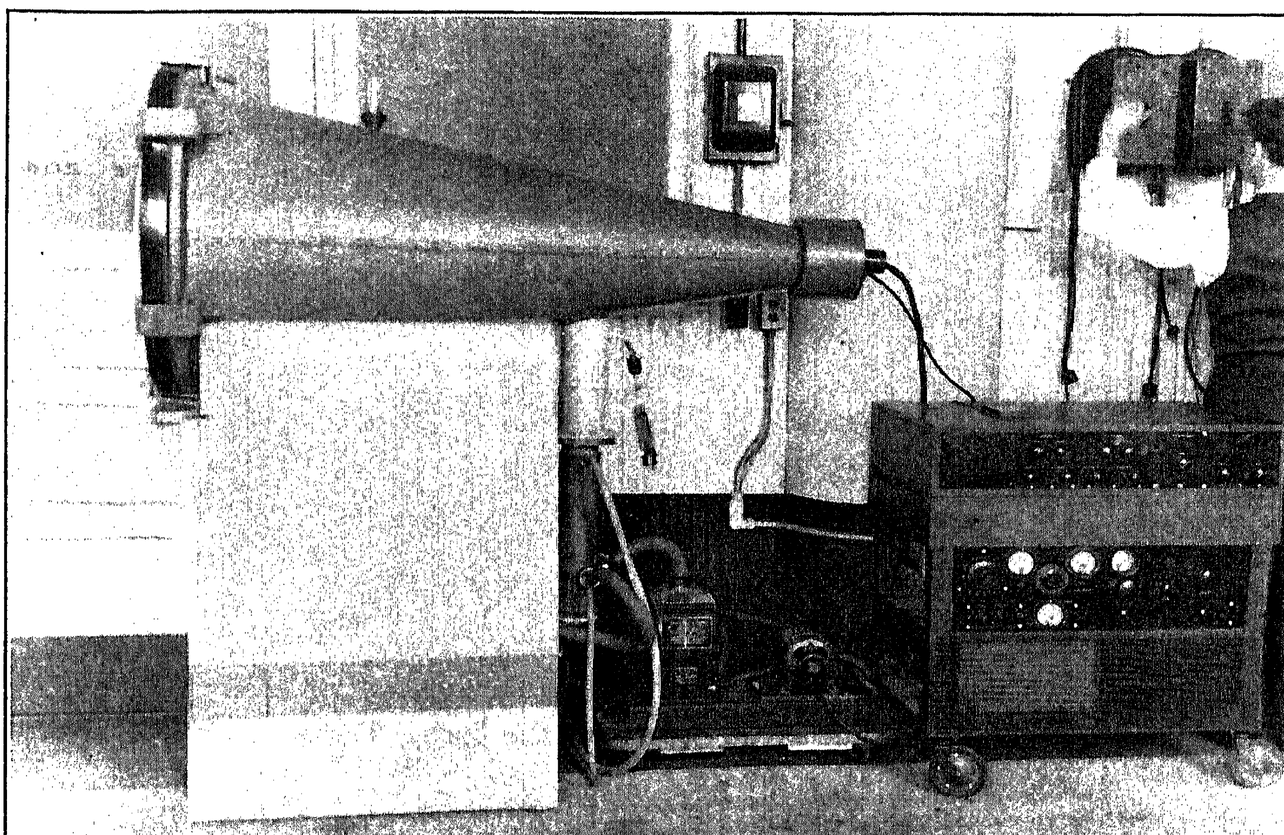


Fig. 12.27.—Thirty-One-Inch Demountable Kinescope.

with the gun, it is usually operated with a 2-milliampere maximum which gives ample brightness. The large end of the cone is closed with a Pyrex disk 31 inches in diameter and 2 inches thick. This thickness is necessary

* See Maloff, reference 12.

in order to withstand atmospheric pressure which results in a force of some $5\frac{1}{2}$ tons on the glass. The fluorescent screen is not the Pyrex disk itself, but instead a thin flat glass sheet coated with the phosphor and mounted well behind the end disk. Fig. 12.27 shows a side view of the tube, including also the pump, power supply, and amplifier. An unre-

FIG. 12.28.—Picture Reproduced on Screen of the Large Demountable Kinescope

touched photograph of a picture reproduced is shown in Fig. 12.28. The excellence of contrast and definition is immediately apparent.

12.11. The Projection Kinescope. The projection Kinescope, as the name implies, is one which produces a picture that is bright enough so that it can be optically projected onto a large viewing screen. Such a tube differs from the direct-viewing Kinescope in two respects: namely, the picture should, for optical reasons, be small, and the peak brightness must be many times greater.

The size of the image on the screen of the projection Kinescope is chiefly determined by the optical system to be used. This, in turn, is dictated by the economics of production. The present indications are that a lens with a 3-inch diameter, an aperture of $f/1.5$, and an angle of field of about 35° most nearly meets the requirements. The size of the image suitable for projection with this lens is about 1.7 by 2.3 inches. This figure should be considered purely tentative and may be radically altered by future developments. Actually, experimental projection tubes are in use having screen diameters ranging from more than 7 inches down to 1 inch. For purposes of the present discussion the figures given above will be accepted.

The illumination required on the viewing screen can be estimated from motion-picture practice. Estimates have been made which show that a highlight brightness of 11 foot-lamberts is required if eye fatigue is to be reduced to an absolute minimum. However, the actual levels attained in moving-picture theaters range from 1 to 9 foot-lamberts, centering about the values given in Table 12.1. As a temporary measure, a brightness of 3.7 foot-lamberts for a 35-mm film, and 2.7 for 16-mm home projectors, has been recommended.

If the figure 2.7 foot-lamberts is assumed for a 1½ by 2 foot picture projected by a tube having a 1.7 by 2.3 inch screen, the surface brightness of this screen must be about 400 candles per square foot, corresponding to a light output of approximately 10 candles. This assumes a directional viewing screen giving nearly five times as much light in a normal direction as a perfectly diffusing screen and a 50 per cent light loss in the lens. If the efficiency of the phosphor is 1½ candles per watt and the overall voltage is 10,000 volts, a beam current of ¾ milliampere will be needed to obtain this brightness. The spot size for this projection tube can be calculated from the dimensions of the picture. It should be no greater than 1.7/441, or about 0.004 inch. The design of a gun meeting this requirement is the major problem of projection-tube development and one which taxes present facilities to their utmost. Guns suitable for projection-tube work will be considered in some detail in the next chapter.

Originally, the projection tube was merely a scaled-down, ordinary, direct-viewing Kinescope. However, it is now recognized as a development in itself, and with this recognition have come some major advances.

Fig. 12.29 illustrates a developmental model of a projection Kinescope with an electrostatically focused gun. The gun is similar in principle to that in the ordinary Kinescope, but has smaller apertures to reduce the spot size. Special precautions are taken to prevent cold discharge and to avoid electrical breakdown, thus permitting the use of 10,000 to 15,000

volts on the gun. The beam is deflected magnetically in both directions, by means of a yoke fitting over the gun neck. The fluorescent screen material is deposited on the end of the tube, which is in the form of a carefully ground and optically polished disk. The optical arrangement for

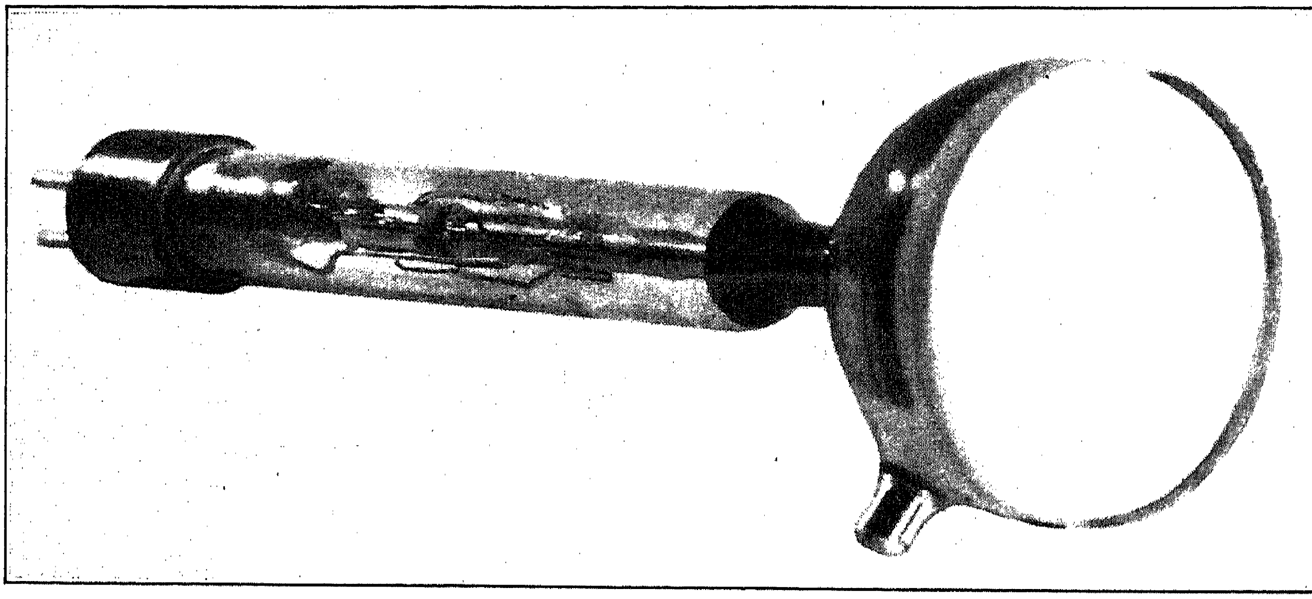


FIG. 12.29.—Early Form of Projection Kinescope.

the tube is shown in Fig. 12.30. A highly directional transmission screen is used.

A second form of projection Kinescope, developed by R. R. Law,* is shown in Fig. 12.31. The gun, which will be described in the next chapter, has a magnetic second lens, together with magnetic deflection. At 10,000 volts a beam current of more than 2 milliamperes delivered into a spot 0.25 mm in diameter can be obtained. This tube, in conjunction with an $f/1.4$ lens, is capable of meeting the 2.7 foot-lambert requirement in a very satisfactory way. Furthermore, a high-definition, high-contrast picture can be obtained.

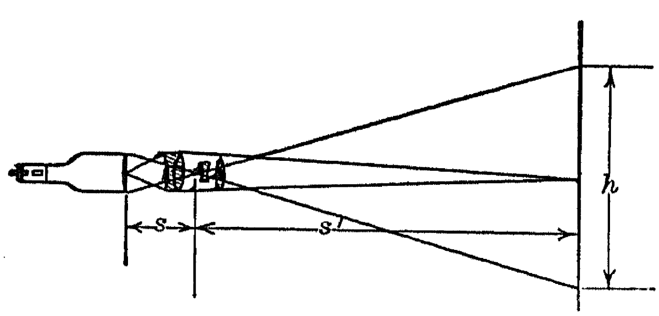


FIG. 12.30.—Optical Arrangement for Projecting Television Pictures.

The two tubes described are merely representative of a large variety of this class. A typical modern projection tube, designed to operate with 20,000 volts, is shown in Fig. 12.32.

In order to increase contrast and also to make more efficient use of the

* See Law, reference 14.

fluorescent material, experiments have been carried out on "front-surface" screens for use in projection tubes. Although interesting results have been obtained with such tubes, they have not been developed to nearly the extent that the above-described tubes have. This is partly

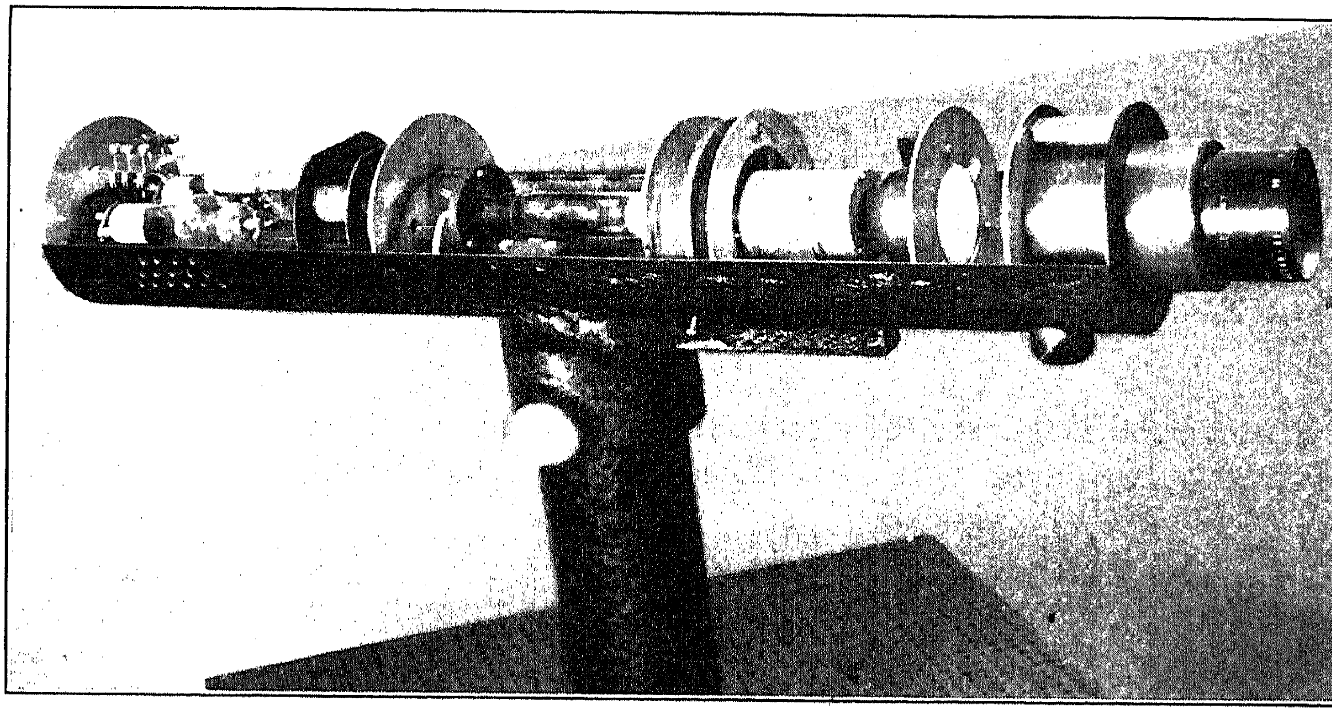


Fig. 12.31.—Modern Television Projector (Law).

because of the unfavorable geometry required for this style of tube, which makes necessary "keystoning" the deflection, tends to result in an elliptical spot due to the angle of the beam, and allows little room for the deflecting yoke. A tube of this type is shown in Fig. 12.33.

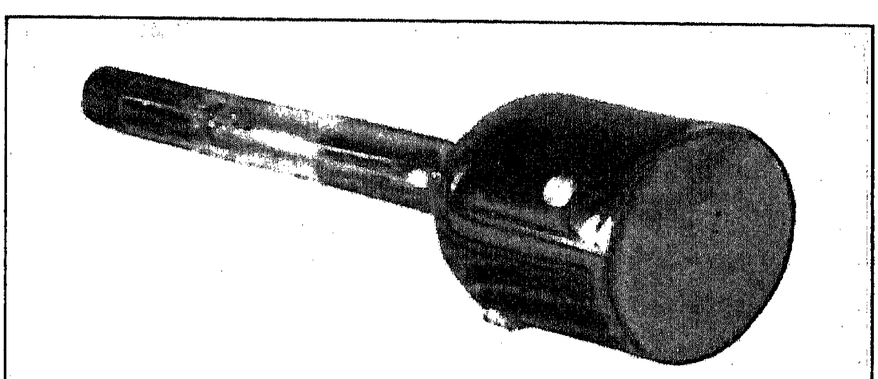


Fig. 12.32.—High-Voltage Projection Kinescope.

Improvements in the optical system to be used with the projection tube have been the subject of considerable investigation. One interesting line of approach has been through the use of a liquid lens. The arrange-

ment is shown in Fig. 12.34. The result of using a lens of this type, with a liquid of index equal to that of the glass, is to increase the light-gathering power of the objective by a factor of 2 to 3 times. Aberrations introduced by the liquid lens, however, have proved difficult to correct. A second, very fruitful, field of optical improvement is through the use of a concave reflector and an aspheric correcting lens instead of the conventional projection lens. Very promising results have already been obtained along this line.

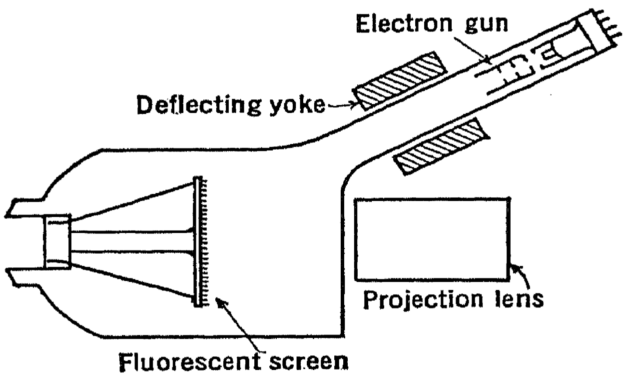


Fig. 12.33.—Front-Surface Projection Kinescope.

Future development of the tube itself probably lies in the direction of greatly increased voltages. This not only introduces new problems in

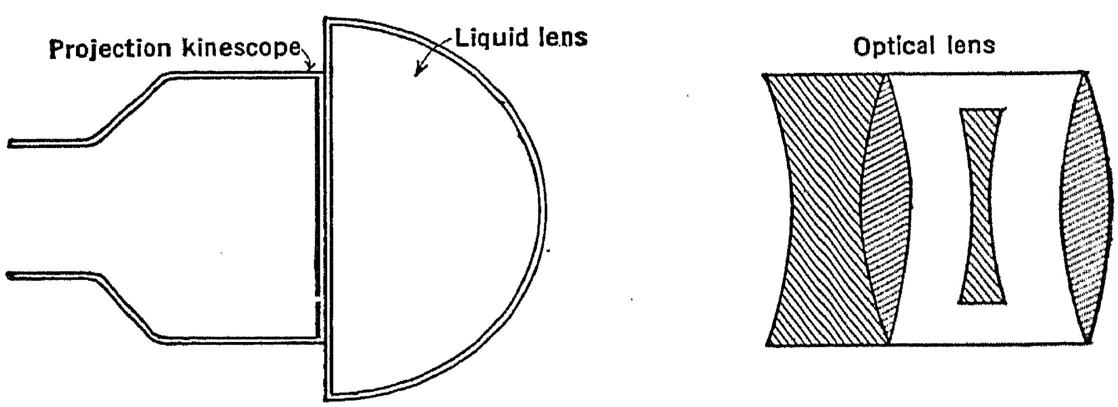


Fig. 12.34.—Liquid Lens to Enhance Brightness of Projected Image.

gun design but also will require further work on the phosphor and methods of applying it.

12.12. Conclusion. At present all practical viewing tubes are based on the use of an electron beam scanning a fluorescent screen. However, much experimental work has been done on radically different types of tubes. These electronic viewing tubes may be divided into three general classes. The first includes those tubes in which the energy of the beam is converted directly into light. Apart from the Kinescope, viewing tubes in which the light for the picture is obtained from a screen of very low heat capacity which is brought to incandescence by the scanning beam are representatives of this group. The second class of tubes includes those in which the electron beam is used to control an electrical effect, which in turn produces light by some form of luminescence. Various mosaic grid image amplifiers and secondary-emission multipliers fall into this cate-

gory. The final class to be mentioned is that in which the scanning beam actuates some form of light valve or shutter.

It is not impossible that, in the future, the last two classes may become of greater importance than the first, because they remove from the beam the burden of supplying the large amount of power necessary to generate the light required for a picture of large dimensions. These investigations must not be ignored, but, as they have not yet been advanced far enough to be considered part of a practical television program, they will not be further discussed in these pages.

REFERENCES

1. Fritz Schröter, Ed., "Fernsehen," Julius Springer, Berlin, 1937.
2. I. G. Maloff and D. W. Epstein, "Electron Optics in Television," McGraw-Hill, New York, 1938.
3. V. K. Zworykin, "Description of an Experimental Television System and the Kinescope," *Proc. I. R. E.*, Vol. 21, pp. 1655–1673, December, 1933.
4. E. W. Engstrom, "A Study of Television Image Characteristics," *Proc. I. R. E.*, Vol. 21, pp. 1631–1651, December, 1933.
5. Max Knoll, "Electron Optics in Television," *Z. tech. Physik.*, Vol. 17, pp. 604–617, 1936.
6. I. G. Maloff, "The Cathode Ray Tube in Television Reception," *Proc. of Radio Club of America*, Oct. 1935, Vol. 12, pp. 31–36. See *Television*, RCA Institutes Tech. Press, Vol. 1, pp. 337–354, July, 1936.
7. V. K. Zworykin, "Iconoscopes and Kinescopes in Television," *R. C. A. Rev.*, Vol. 1, pp. 60–84, July, 1936.
8. H. W. Leverenz, "Problems Concerning the Production of Cathode Ray Tube Screens," *J. O. S. A.*, Vol. 27, pp. 25–35, January, 1937.
9. C. E. Burnett, "A Circuit for Studying Kinescope Resolution," *Proc. I. R. E.*, Vol. 25, pp. 992–1011, August, 1937.
10. R. R. Law, "Contrast in Kinescopes," *Proc. I. R. E.*, Vol. 27, pp. 511–524, August, 1939.
11. R. S. Burnap, "Television Cathode Ray Tubes for the Amateur," *R. C. A. Rev.*, Vol. 2, pp. 297–302, January 1938.
12. I. G. Maloff, "Direct Viewing Type Cathode Ray Tube for Large Television Images," *R. C. A. Rev.*, Vol. 2, pp. 289–296, January, 1938.
13. V. K. Zworykin and W. H. Painter, "Development of the Projection Kinescope," *Proc. I. R. E.*, Vol. 25, pp. 937–953, August, 1937.
14. R. R. Law, "High Current Electron Gun for Projection Kinescopes," *Proc. I. R. E.*, Vol. 25, pp. 954–976, August, 1937.

CHAPTER 13

THE ELECTRON GUN

The concentration of electrons into a narrow beam is one of the very early phenomena observed in connection with cathode rays. Even before the nature of cathode rays was understood, it was known that in a gas-discharge tube the rays could be focused into a narrow pencil by properly shaping the disk from which the discharge originates. For example, this property was applied in the early gas X-ray tubes by Roentgen and others.

These early experiments also showed that such a beam could be directed onto a fluorescent screen, and deflected by means of an electrostatic or magnetic field. The first use of these effects was solely for the determination of the properties of cathode rays and of the electron. However, the recognition that the deflection of the focused spot could be used to measure the fields soon led to the development of the cathode-ray oscilloscope. Experiments by Rosing, which have already been described, used the gas-discharge oscilloscope as a means of reproducing television images. The last-mentioned field of application of the focused cathode-ray beam, namely television, has gained an importance which is rapidly increasing. The requirements placed by this application on the means for producing the electron pencil, which is usually termed the electron gun, are very severe. In consequence, the design of the electron gun, for either a pickup tube or a viewing tube, has become a major problem.

The earliest cathode-ray tubes were gas-discharge tubes in which the low-pressure gases served not only to concentrate the beam but also to generate the electrons. The difficulty of controlling the current in the beam, and the fact that a beam focused in this way cannot be deflected at high speeds, made these tubes inadequate.

The first-mentioned difficulty was overcome by the introduction of a thermionic cathode to serve as the electron source. This made it possible to govern the current in the beam with a control grid. Such tubes were found applicable to the reproduction of low-definition television images. However, there still remained the difficulty incidental to deflection which made them unsuited for high-definition pictures. The difficulty arises from the nature of gas focusing, and its dependence upon the presence of a positive-ion core at the center of the beam, which will not follow a

rapid lateral motion of the beam. So fundamental did this obstacle prove to be that it became necessary to abandon the gas tubes in favor of hard cathode-ray tubes.

Without gas to aid the focusing of the beam, the design of the electron gun is much more difficult. The earliest hard cathode-ray tubes made use of an electrode system which produced a concentrating electrostatic field in the immediate neighborhood of the cathode; this field was intended to cause the electrons from the source to converge upon a point on the viewing screen. An example of this type of concentrating system is to be found in the Wehnelt cylinder illustrated in Fig. 13.1. Such relatively simple electron-optical arrangements were satisfactory as long as the picture requirements did not demand a very small spot or very high current densities. However, when these two conditions became a necessity, a more complicated gun was required.

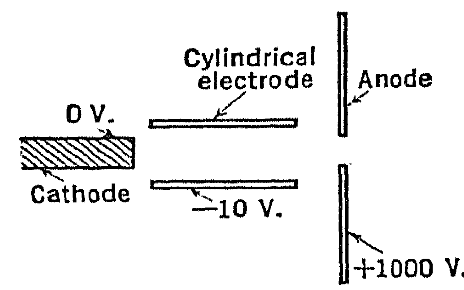


Fig. 13.1.—Electron Gun with Wehnelt Cylinder.

13.1. Requirements of the Electron Gun. Before continuing with the discussion of the gun itself, the characteristics required of it should be reviewed.

Electron guns may be divided into three classes, depending upon the use to which they are to be put, namely, guns for the pickup tube, for the direct-viewing tube, and for the projection tube.

The Iconoscope, or pickup tube, gun operates at relatively low power, but must be capable of forming an extremely small, well-defined spot. For this purpose the beam electrons are rarely required to have velocities in excess of a thousand volts. Beam currents of 0.1 to 0.5 microampere are generally used, although certain special mosaics may require much higher currents. Although it is necessary that the beam current be controllable, the control characteristics are relatively unimportant. The spot must have a diameter not much greater than 0.01 cm for mosaics of the ordinary size; the meaning of the term spot diameter will be given later. It is important that the spot shape be regular and the spot well defined.

The requirements for guns used in viewing tubes are very different. Bombarding velocities up to 6000 volts, or even higher, are necessary for efficient utilization of the fluorescent screen. The current required in the beam ranges from 100 to 500 microamperes, depending upon the type of tube. The spot size depends upon the size of the viewing screen, but is always much greater than that for the Iconoscope. Where a 12-inch screen is employed, a spot diameter of 0.07 cm will be adequate.

The gun in a projection Kinescope must be capable of forming a very small spot carrying a high current. The voltages used may be 20,000 volts or higher, and currents of 2 or 3 milliamperes are not infrequent. Again the spot size depends upon the screen size; for example, where a 2 by 2.6-inch picture is to be formed, the spot diameter must be of the order of 0.025 cm. As can be readily appreciated, these requirements are very severe and can be met only with difficulty, if at all.

Guns used in both types of viewing tubes must be designed in such a way that the current in the scanning beam can be readily controlled by the output from the video amplifier. Moderate voltage swing of the control element, low power requirements, and linearity of control are all desirable features.

These requirements are listed only for purposes of orientation; many circumstances which are not at all exceptional may demand very different operating conditions.

13.2. Basis of Electron-Gun Design. All, or nearly all, electron guns used at present for television purposes are based on a two-lens principle. Such guns consist of a thermionic electron source, a first lens, almost invariably electrostatic, and a second lens which may be either electrostatic or magnetic, or a combination of the two.

The first lens is of the cathode lens type described in Chapter 4 and serves to accelerate the electrons away from the cathode and to converge them into a bundle of narrow cross-section. This narrow portion of the ray bundle is not an image of the emitter, but corresponds more nearly to the exit pupil of the first lens system, and is usually referred to as the "crossover." Associated with the first lens is the control system. Various means of control have been successful, such as electrodes producing fields which restrict the emitting area of the cathode, elements causing the beam to be deflected or defocused at an aperture, and controls based on ordinary space-charge principles. Combinations of two or more of these methods are frequent.

The second lens images the crossover formed by the first lens onto the mosaic or fluorescent screen. With deflection, or defocusing, control, the second lens may image the restricting aperture rather than the crossover when these two do not coincide. In general, however, imaging the crossover itself leads to the maximum current for a given spot size. Where an electrostatic system is used for the second lens, it is ordinarily arranged to have a voltage ratio greater than unity so that the beam electrons are accelerated by the lens. In the case of a magnetic second lens the electrons receive their full acceleration at the first lens. Each of these lens systems has advantages and disadvantages which will become apparent as the discussion proceeds.

13.3. Current in an Ideal Gun. The primary consideration in electron-gun design is to obtain a spot of small diameter and high current density. It is, therefore, important to investigate the factors upon which these two properties depend. The analysis of an actual gun is very complicated and, in general, cannot be rigorously carried out. However, by postulating an idealized gun,* it is possible to determine the dependence of the current and spot size upon at least some of the parameters of the system.

The ideal gun will be assumed to consist of a cathode and a first and second lens system. The control grid, usually part of the first lens system,

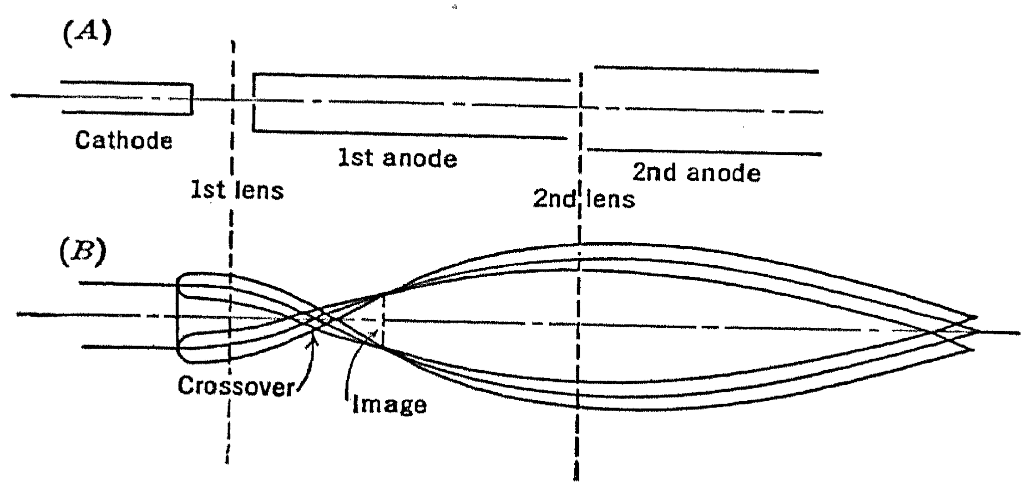


FIG. 13.2.—Electrode Arrangement and Electron Ray Paths in Ideal Gun.

will be ignored entirely, no account will be taken of space charge, and, furthermore, both lens systems will be assumed to be free from aberrations.

Fig. 13.2*a* shows the general arrangement of the ideal gun; a schematic representation of the electron paths through the gun is shown in Fig. 13.2*b*. Electrons are emitted from every point of the cathode in all directions, as dictated by the distribution laws of thermionic emission. These electrons are accelerated toward, and through, the lens. The principal electron rays from all points of the cathode converge at the second focal point of the first lens. Other electrons with radial initial velocities are, of course, displaced a small radial distance from the axis at this point. Beyond this point the electron ray bundle from each point on the emitter may converge, forming a real image of the cathode, or the electrons of each bundle may diverge and the image of the cathode be virtual. In either case the diameter of the group of electrons is a minimum at the point where the principal rays cross the axis. It has been noted that this point, which in the electron gun is designated as the crossover, is analo-

* See also Langmuir, reference 11; and Law, reference 12.

gous to the exit pupil in conventional optics. The second lens, when adjusted to give a minimum spot instead of using the image of the cathode formed by the first lens as its object, images the crossover on the screen.

The electrons are emitted from the thermionic cathode with initial velocities which have a Maxwellian distribution. If the radial velocities are expressed in electron volts, the number of electrons having radial initial velocities lying between ϕ_r and $\phi_r + d\phi_r$ will be:

$$B_{\phi_r} d\phi_r = A\epsilon^{\frac{-e\phi_r}{kT}} d\phi_r \tag{13.1}$$

where A is a constant of emission, k is Boltzmann's constant, and T is the temperature of the cathode in degrees absolute. On this basis, the current I_V, consisting of electrons having radial velocities less than, or equal to, V volts, is given by:

$$I_V = e\int_0^V B_{\phi_r} d\phi_r. \tag{13.2}$$

For purposes of this analysis, it is convenient to assume that the first lens is arranged as indicated in Fig. 13.3. The electrons are accelerated

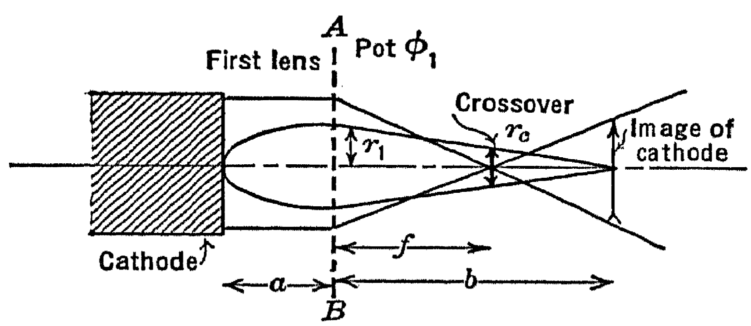


Fig. 13.3.—First Lens of Ideal Gun.

by a uniform field up to the lens, which is indicated by the dotted line AB. The lens, which has a focal length f, bends the electrons toward the axis. A potential ϕ_1 will be assumed at the lens, and, furthermore, the region beyond the first lens up to, and including, its second focal point, will be considered field free.

An electron emitted from the cathode at a point on the axis with an initial radial velocity ϕ_r will follow a parabolic path until it reaches the lens AB. At the lens its radial distance from the axis becomes:

$$r_1 = 2\sqrt{\frac{\phi_r}{\phi_1}}a, \tag{13.3}$$

where a is the distance between the lens and the cathode. The lens will deflect the electron toward the image point corresponding to the emitting

point. If the second focal length is f, the distance b between the image and the lens is given by the relation:

$$\frac{1}{2a} + \frac{1}{b} = \frac{1}{f}. \tag{13.4}$$

The radial distance r_c between the electron and the axis at the crossover (i.e., the second focal point) can be found as follows:

$$\begin{aligned} r_c &= r_1 \frac{b-f}{b} \\ &= 2\sqrt{\frac{\phi_r}{\phi_1}}\, a \frac{b-f}{b} \\ &= \sqrt{\frac{\phi_r}{\phi_1}}\, f. \end{aligned} \tag{13.5}$$

It has been assumed that the electrons had their origin on the axis. For electrons from a point off the axis, the principal rays cut the axis at the crossover as before, but the intersection of the plane of the crossover and the ray bundle will be an ellipse whose minor axis is the radius r_c of the ray bundle and whose major axis is equal to $r_c/\cos\alpha$, where α is the angle between the axis and the principal ray. In all practical cases α is small, so that $\cos\alpha$ can be assumed to be unity, and Eq. 13.5 can be used in the consideration of an extended cathode.

The radial initial velocities and the radius of the circle of electrons at the crossover are obviously related by virtue of Eq. 13.5. With the aid of Eqs. 13.1 and 13.2, the fraction of the total emitted current having initial velocities no greater than ϕ_r can be determined and is given by:

$$\frac{I_c}{I_T} = \frac{\int_0^{\phi_r} \epsilon^{\frac{-e\phi_r}{kT}}\, d\phi_r}{\int_0^{\infty} \epsilon^{\frac{-e\phi_r}{kT}}\, d\phi_r}.$$

Carrying out the indicated integration, this relation becomes:

$$\frac{I_c}{I_T} = 1 - \epsilon^{\frac{-e\phi_r}{kT}} \tag{13.6}$$

By substitution from Eq. 13.5 in Eq. 13.6, the current within a radius r_c at the crossover can be written as:

$$I_c = I_T\left(1 - \epsilon^{\frac{-r_c^2}{f^2}\frac{e\phi_1}{kT}}\right). \tag{13.7}$$

Furthermore, by differentiating with respect to r_c and dividing by $2\pi r_c$,

the current density as a function of r_c can be found. It is given by the following relation:

$$\rho(r_c) = \frac{I_T}{\pi f^2} \frac{e\phi_1}{kT} \epsilon^{\frac{-r_c^2}{f^2} \frac{e\phi_1}{kT}}. \tag{13.8}$$

Since the second lens is adjusted to image the crossover onto the viewing screen, or mosaic, any circle of radius r_c will be imaged into a circle of radius r_s at the spot. The second lens is shown schematically in Fig. 13.4 as a thin lens represented by the line CD at distances u and v from the crossover and screen, respectively. The potential in the image field of the

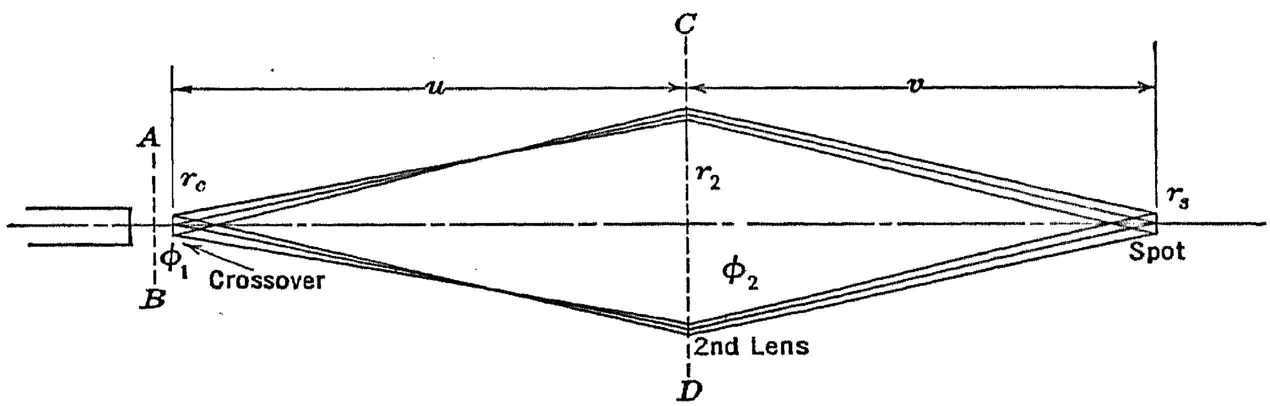


FIG. 13.4.—Second Lens of Ideal Gun.

second lens will be taken as ϕ_2. Therefore, the radii of the image and object will be related by:

$$\frac{r_s}{r_c} = \frac{v}{u}\sqrt{\frac{\phi_1}{\phi_2}}. \tag{13.9}$$

The fraction of the total emitted current within a radius r_s of the spot, and the current density distribution, can be found with the aid of Eq. 13.9 from Eq. 13.7. These relations are:

$$\frac{I_s}{I_T} = 1 - \epsilon^{\frac{-r_s^2 e\phi_2}{f^2\left(\frac{v}{u}\right)^2 kT}} \tag{13.10}$$

and:

$$\rho(r_s) = \frac{I_T}{\pi f^2\left(\frac{v}{u}\right)^2} \frac{e\phi_2}{kT} \epsilon^{\frac{-r_s^2}{f^2\left(\frac{v}{u}\right)^2} \frac{e\phi_2}{kT}} \tag{13.11}$$

Next, the diameter of the beam can be determined as it passes through the second lens. For the aberration-free lens which has been assumed this is of no consequence, but for an actual lens in which spherical aberration

plays an important role, the spread of the beam must be considered. Fig. 13.5 again shows the idealized gun, with principal rays originating at the edge of the cathode a distance r_c off the axis. These, after passing

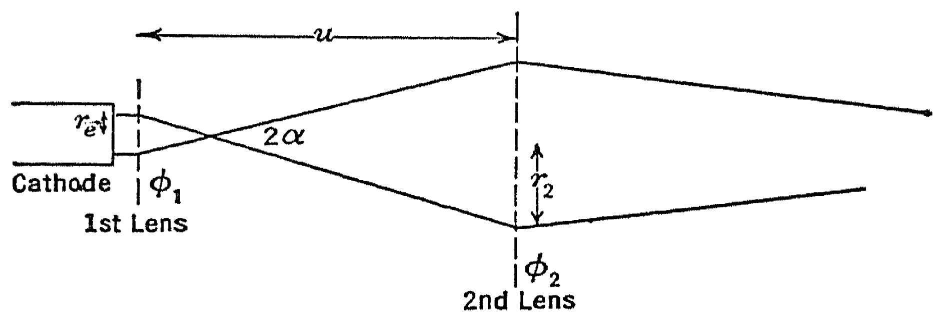


FIG. 13.5.—Marginal Rays through Second Lens.

through the first lens, are deflected so as to meet the axis at the crossover. At this point they make an angle to the axis of:

$$\alpha \cong \frac{r_c}{f}$$

and have an axial velocity:

$$v_z \cong \sqrt{\frac{2e\phi_1}{m}}. \tag{13.12a}$$

Hence, their radial component of velocity is:

$$v_r = \frac{r_c}{f}\sqrt{\frac{2e\phi_1}{m}}. \tag{13.12b}$$

If, now, the second lens is a true thin lens, and the region between it and the crossover is essentially field free, the electrons will cross the intervening space at constant velocity v_z along straight lines. At the second lens nearly all the electrons will lie within a circle formed by those electrons which have a radial velocity v_r. The radius of this boundary is:

$$r_2 = \frac{v_r}{v_z}u = \frac{r_c u}{f}. \tag{13.13}$$

Certain important conclusions can be drawn from Eqs. 13.7, 13.8, 13.10, 13.11, and 13.13 concerning the operating characteristics of the gun. Because of the simplifying assumptions made as to the nature of the first and second lens, caution is necessary in applying these deductions to an actual gun.

Eq. 13.11, which describes the current distribution in the spot, indicates that the spot has no sharp boundary but instead the current density decreases exponentially with the square of the distance from the center.

Therefore, the term "spot size" must be arbitrarily defined. One very practical definition, which has operational significance, is that which limits the radius of the spot to a value for which the current density is reduced to a certain fraction of the density at the center, or, symbolically:

$$K = \frac{\rho(\beta)}{\rho(0)} = \epsilon^{\frac{-\beta^2}{f^2\left(\frac{v}{u}\right)^2}\frac{e\phi_2}{kT}}, \tag{13.14}$$

where β is the radius of the spot and K the limiting fraction of current density. The actual value of K is unimportant for purposes of the present discussion. In practice it depends upon the use to which the gun is to be put.

Eq. 13.14 can be rewritten as:

$$\beta = f\left(\frac{v}{u}\right)\sqrt{\frac{-kT}{e\phi_2}\ln K} \tag{13.15}$$

$$= \frac{F_1F_2}{\sqrt{\phi_2}}G, \tag{13.15a}$$

F_1 and F_2 being functions of the properties of the first and second lens, respectively, and G a constant which depends on the density ratio.

It is evident from Eq. 13.15a that the "spot size" is independent of the first lens voltage ϕ_1; furthermore, it is seen to be inversely proportional to the square root of the total voltage ϕ_2.

Eq. 13.10, together with Eq. 13.14, can be used to determine the current in the spot, as follows:

$$\begin{aligned} I_s &= I_T\left(1 - \epsilon^{\frac{-\beta^2 e\phi_2}{f^2\left(\frac{v}{u}\right)^2 kT}}\right) \\ &= I_T(1 - K). \end{aligned} \tag{13.16}$$

In other words, practically all the current emitted by the cathode is concentrated into the spot. This result is to be expected since, as yet, no account has been taken of possible limiting apertures.

The radius of the beam at the crossover (defined in the same terms as the spot size) varies inversely with the square root of the first lens voltage. As has been shown, in the absence of limiting apertures, the size of the spot is independent of the crossover diameter. However, this does not mean that the first lens voltage is unimportant. In the first place it should be high enough to saturate the emission from the cathode when the maximum beam current is required. Furthermore, it must be large enough to minimize the defocusing action of the space charge in the

neighborhood of the cathode. On the other hand, a high first lens voltage tends to make the control of the electrons more difficult.

In the idealized gun postulated, the first lens voltage has no effect on the diameter of the beam through the second lens. If a potential gradient had been assumed between the first and second lens the diameter of the ray bundle at the second lens would have been found to vary with the first lens voltage. Measurement on representative practical guns shows almost no variation in the diameter of a cathode-ray beam at the second lens as the first lens voltage is changed.

The radius of the ray bundle through the lens is also directly proportional to the radius of the emitting area. When the current control for the beam is obtained by restricting the emitting area of the cathode, the size of the beam at the second lens varies with beam current. Under these conditions, the presence of spherical aberration in this lens will cause the spot size, also, to vary with beam current.

Because of the aberrations inherent in all electron lenses, the spot size in a real gun will be considerably greater than that of the ideal gun. In order to minimize the effect of these aberrations, it is common practice to distribute limiting apertures throughout the system.

The most important positions for the limiting apertures are: at the cathode, at the crossover, and at the second lens. In a practical gun, the apertures cannot, in general, be placed exactly in these positions for reasons connected with the physical construction of the lenses. However, no such restriction enters into the consideration of the ideal gun.

The effect of restricting the radius of the emitter leads to no first-order change in the spot size. However, for a given specific emissivity of the cathode material, the beam current is directly proportional to the area of the cathode, i.e., to the square of the radius. The diameter of the beam through both the first and second lenses is approximately proportional to the diameter of the cathode. Because of this, for reasons connected with the aberrations of these lenses, in particular of the first lens, a very large cathode area cannot be used without increasing the spot size. By curving the surface of the emitter some of these defects can be minimized, thus permitting a larger cathode area and, consequently, a higher beam current.

An aperture at the crossover has many consequences. The spot size under these conditions is no longer independent of the first lens potential but is given by the expression:

$$r_s = \beta = r_a \frac{v}{u} \sqrt{\frac{\phi_1}{\phi_2}}, \tag{13.17}$$

where r_a is the radius of the aperture. The current passing through the aperture can be expressed by the integral:

$$\begin{aligned} I_b &= \int_0^{r_a} 2\pi r \rho_c(r)\,dr \\ &= I_T \int_0^{r_a} \frac{r}{f^2}\frac{2e\phi_1}{kT}\epsilon^{\frac{-r^2}{f^2}\frac{e\phi_1}{kT}}dr \\ &= I_T(1 - \epsilon^{\frac{-r_a{}^2}{f^2}\frac{e\phi_1}{kT}}). \end{aligned} \tag{13.18}$$

These two relations indicate that the spot size can be decreased by decreasing the radius of the crossover aperture or by reducing the first anode voltage. However, this reduction can be accomplished only at the expense of beam current. It should be noticed that the two methods of reducing the spot size cause an equal decrease in beam current. Both have little effect on the diameter of the beam at the second lens.

When the second lens is provided with an aperture to limit the beam diameter, its effect on the aberration-free system is practically the same as limiting the area of the emitter. The difference in action is due to the fact that the diameter of the beam at this lens is not exclusively determined by the marginal principal rays, as was heretofore assumed, but is also a function of the crossover diameter. In a practical gun, an aperture placed at, or near, this point, to reduce spherical aberration of the lens, limits the spread of the beam resulting from the aberrations in the first lens, as well as that due to the first-order paths of the rays, and thus is very much more effective for the purpose than limiting the cathode area (although the cathode is usually restricted as well to avoid excessive loss of current to the aperture limiting the second lens).

A further analysis of the ideal gun is interesting academically, but not of sufficient practical importance to warrant inclusion in this discussion. The possibility of using the image of the cathode formed by the first lens as object for the second lens should not be overlooked by anyone interested in gun design. An analysis following the lines of the preceding treatment does not present any particular difficulties.

13.4. The Cathode. The design of the thermionic cathode and the choice of emitting material for the practical gun are of fundamental importance. In the preceding section it was pointed out that it is generally necessary to restrict the emitting area, and, consequently, the current which can be obtained in a spot of given size is, except for space-charge limitations, directly proportional to the emissivity of the material. Furthermore, it was shown that the radial initial velocities play an important role in determining the distribution of current at the crossover, and consequently at the spot. Another consequence of high initial velocities,

which was not mentioned in the discussion of the ideal gun, is that it increases the spot size due to chromatic aberration in the first lens. In order to minimize these undesirable effects, the cathode should be operated at as low a temperature as possible, as is evident from the form of the thermionic emission equations.

The velocity distribution of the electrons from the various cathodes can be determined from the Maxwell-Boltzmann distribution law. According to Eq. 13.1 the relation giving the number of electrons with radial velocities lying between ϕ_r and $\phi_r + d\phi_r$ electron volts is:

$$Bd\phi_r = A\epsilon^{\frac{-e\phi_r}{kT}} d\phi_r.$$

Similarly, the axial distribution is given by:

$$Dd\phi_a = E\epsilon^{\frac{-e\phi_a}{kT}} d\phi_a.$$

These two initial velocity distributions are important in determining the crossover radius and chromatic aberration of the first lens.

Cathodes involving a caesium activation would be highly desirable from the standpoint of high emissivity and low operating temperatures. However, these materials are, in general, much too delicate for this purpose, and are de-activated almost immediately by ion bombardment.

Both thoriated tungsten and cathodes coated with barium and strontium oxide are suitable for use in electron guns. Owing to their high specific emissivity and lower operating temperature, the oxide-coated cathodes are preferable, and are used almost universally in sealed-off tubes. Thoriated tungsten cathodes are often advantageous in demountable systems because this type of emitter can be readily reactivated.

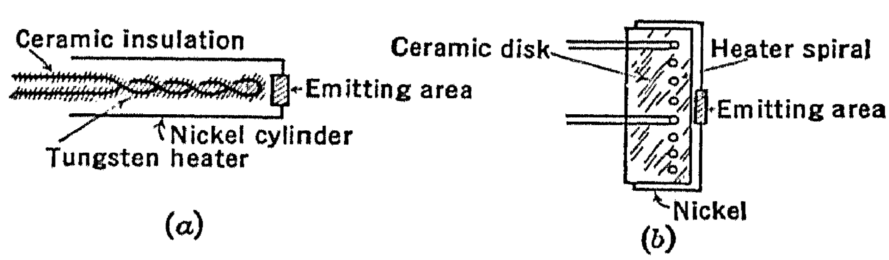


Fig. 13.6.—Typical Indirectly Heated Cathodes.

Where more rugged cathodes are required pure tungsten or tantalum may be employed.

The conventional oxide-coated cathode assembly consists of a heating element, a cathode cup, and the emitter. Indirect heating of the cathode has the advantage of minimizing the interaction between the magnetic and electrostatic fields produced by the heater current and the beam

electrons. Typical forms of oxide-coated emitters are illustrated in Fig. 13.6. The heater is usually a tungsten filament covered with a refractory material to insulate it, and fitted inside a nickel cup. The nickel cup serves as a base for the cathode material. The cathode area of the nickel cup is coated with a mixture of barium and strontium carbonate. This material can be decomposed by heating and activated as described in Chapter 1.

The cathode used in the typical Kinescope is assembled as shown in Fig. 13.7. The heater is a tungsten-twist designed to operate at 2.1 amperes and 2.5 volts. This double spiral is insulated from itself and from the cathode cup with a refractory clay. A cylinder closed at one end with a recessed disk forms the cathode cup. The heater is inserted in the cathode cup, which it heats partly by radiation and partly by conduction. The emitting material just fills the hollow in the end of the cup so that the recess serves to hold the material in place and to limit the emitting area.

A curved cathode is sometimes advantageous to minimize the aberrations in the first lens. A cathode of this type, used in the gun of the Law projection tube, will be illustrated in section 13.9.

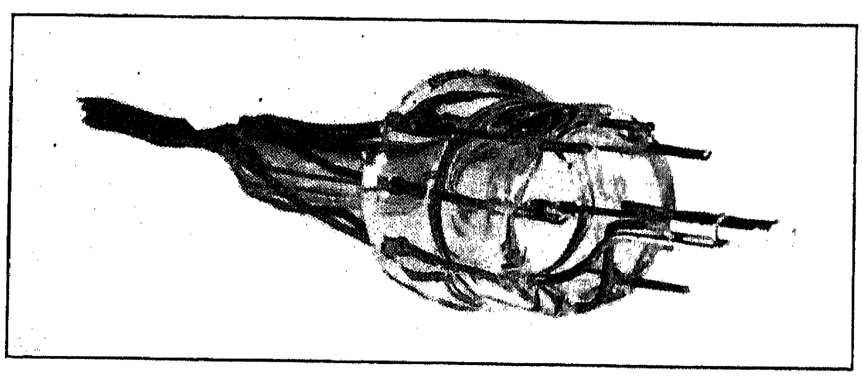


Fig. 13.7.—Photograph of Cathode Assembly.

The assembly of a thoriated tungsten cathode depends upon the kind of demountable vacuum tube for which it is intended. Usually, the cathode is directly heated and is in the form of a thoriated tungsten ribbon or wire. In order to restrict the emitting area, the cathode is covered

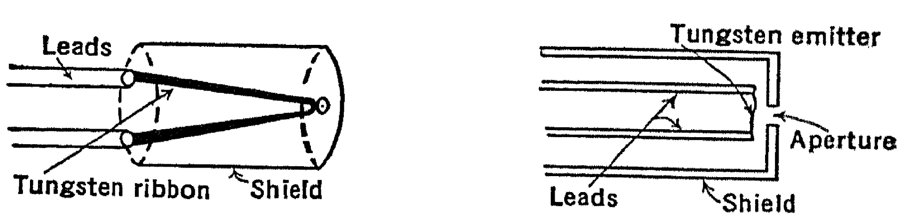


Fig. 13.8.—Two Forms of Demountable Cathodes.

with an apertured shield. Two typical cathode assemblies of this kind are shown in Fig. 13.8.

13.5. First Lens and Control Grid. The simplest first lens system is composed of three elements: the cathode, the control grid, and the first anode. The arrangement is illustrated in Fig. 13.9. The control grid consists of a short cylinder closed by an apertured disk close to the end of the cylinder. The fact that the disk is placed so as to allow a portion of the cylinder known as the grid skirt to extend beyond it is important from the standpoint of the control characteristic. The first anode is also a cylinder, of the same diameter as the grid, and is closed at the grid end by an apertured disk.

The first lens region in the electron gun is, from an analytic point of view, by far the most difficult part of the whole electron-optical system.

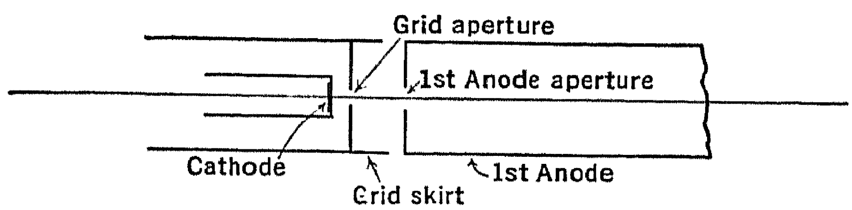


Fig. 13.9.—Simple Three-Element First Lens.

The electrostatic field in this region is so complicated that it is completely unamenable to mathematical manipulation. Furthermore, the control grid, which forms a part of the first lens, varies in potential, altering the optical properties of the system. Finally, the electron densities in this region are sufficiently great so that the space charge appreciably alters the field. In spite of these difficulties, this region has been sufficiently analyzed so that the behavior of the electrons over this portion of their trajectories is fairly well understood.

Fig. 13.10 shows the equipotential surfaces in the neighborhood of the cathode in a typical three-element first lens system for two values of grid potential, as measured by Maloff and Epstein.*

These field maps show that the potential of the control grid acts in three different ways. At, and near, the center of the cathode it controls the current in the beam by forming a space-charge barrier in the manner of a conventional control grid. Furthermore, the control grid acts to restrict the area from which emission can take place. Thus, even if the emission is saturated at the center of the cathode, the current in the beam can be increased by decreasing the negative bias applied to the grid, because this will permit electrons to leave from a greater area of cathode. The final consequence of varying the control grid voltage is to alter the

* See Maloff and Epstein, reference 8.

position of the cardinal points of the lens system. As the grid is made more negative, the second focal point and, consequently, the crossover move closer to the cathode.

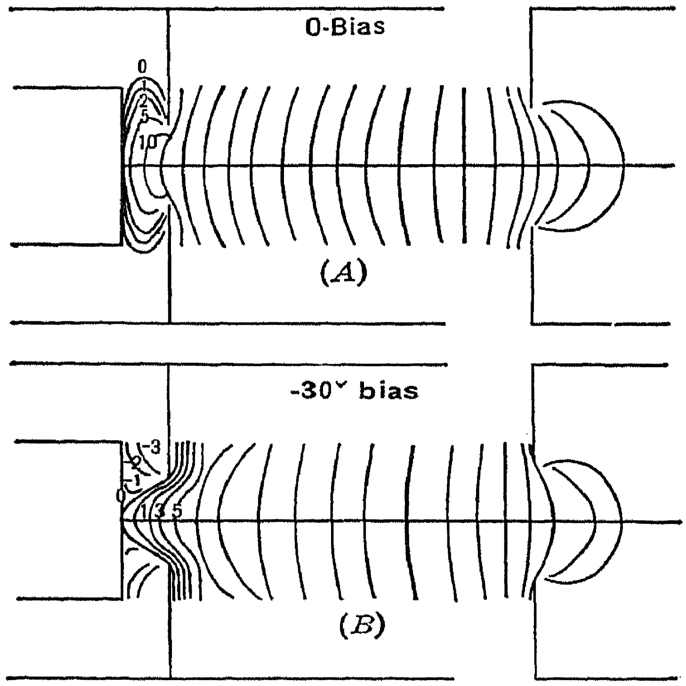


FIG. 13.10.—Equipotentials in Three-Element First Lens (after Maloff and Epstein).

The cardinal points of the first lens have been determined on the basis of potential maps. The general arrangement of these points is shown in Fig. 13.11. Their position depends, of course, upon the velocity with which the electrons are emitted. Calculations show that, in the particular

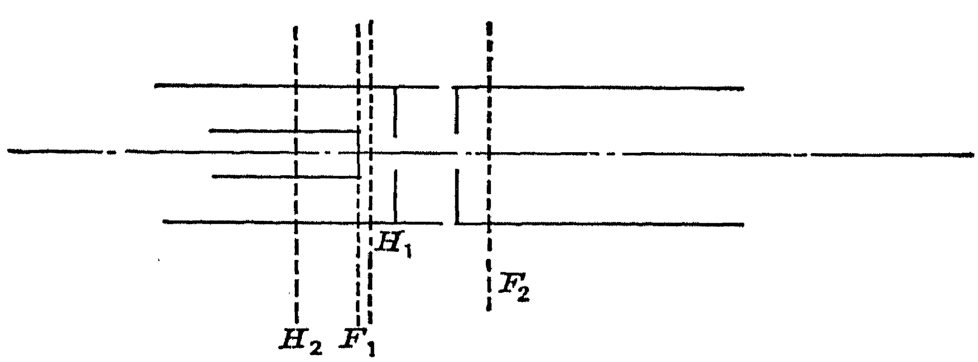


FIG. 13.11.—Cardinal Points of First Lens (after Maloff and Epstein).

configuration with the bias at −30 volts, the second principal point and focal point are rather insensitive to the initial velocities, while the first focal point and principal point move from the surface of the cathode to a distance of a few thousandths of a centimeter in front of the cathode as the velocity goes from zero to 0.5 volt. It is also found that at zero

volt bias the shift of the cardinal points, corresponding to chromatic aberration in the first lens, is much less.

The equipotential plot in Fig. 13.10, determined by means of an electrolytic plotting tank, represents the field configuration in the absence of space charge. This approximation is not sufficiently good for purposes of determining the actual performance. In order to estimate* the effect of space charge, the configuration and charge density of the beam in the absence of space charge are calculated, and from these the new potential is determined. To attain a better approximation this procedure may be repeated. Fig. 13.12 shows the envelope and density distribution of the beam; Fig. 13.13 illustrates the potential distribution corrected for the effect of space charge.

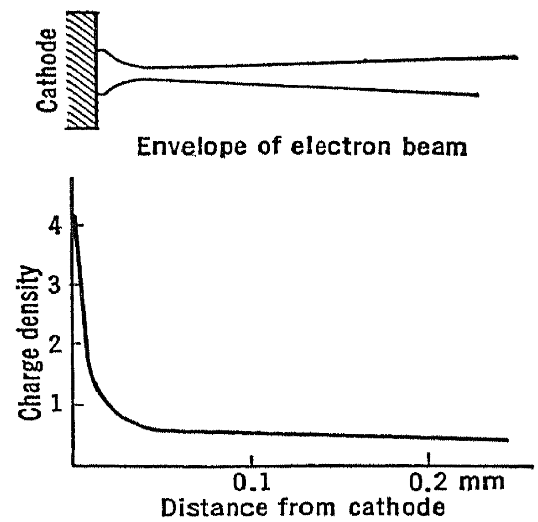


FIG. 13.12.—Envelope and Electron Density of Beam in First Lens (after Maloff and Epstein).

Three important conditions must be satisfied by the ideal first lens system: (1) the configuration must be such as to permit the flow of high current without too large a crossover (i.e., the product of first anode voltage and crossover area must be small); (2) the control grid characteristics must be such as to permit ease of control; and finally, (3) the effect

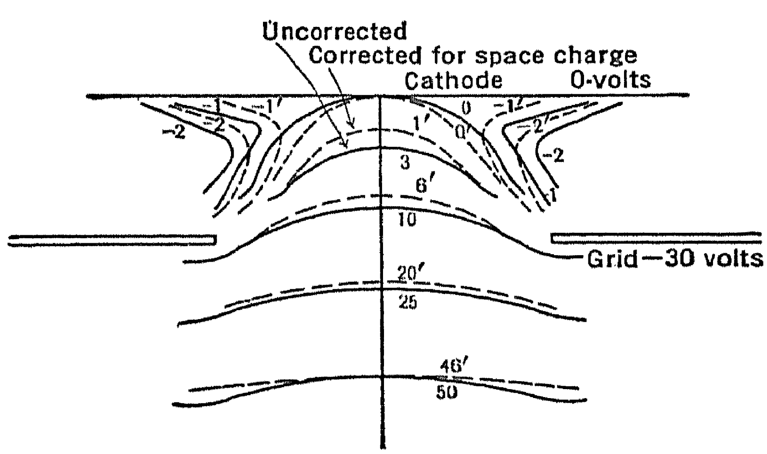


FIG. 13.13.—The Effect of Space Charge in the First-Lens Region (Maloff and Epstein).

of control voltage on spot size must be a minimum. No one first lens design has been worked out which satisfies completely all these conditions. There are several systems, however, which are very good from a practical standpoint.

* See Maloff and Epstein, reference 8.

The system which was described above can, by suitably selecting the size of the grid aperture and the length of the grid skirt, be made to have excellent control characteristics and to deliver considerable current into

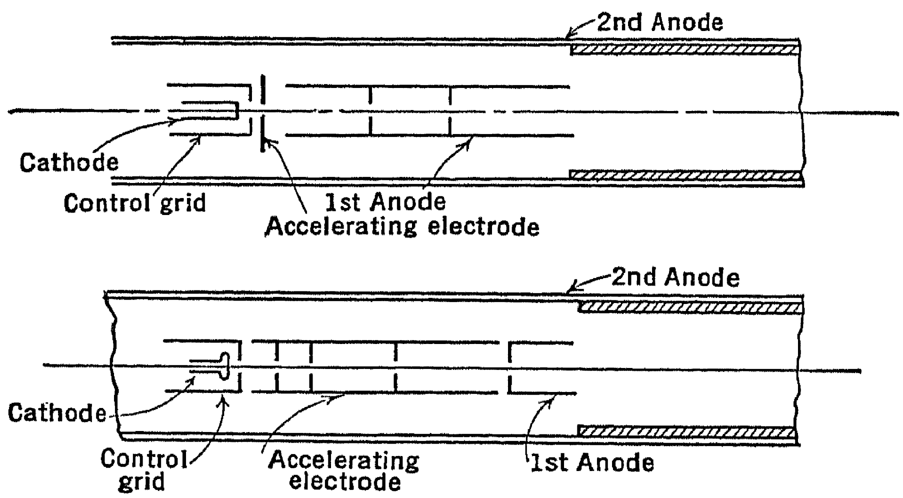


Fig. 13.14.—Lens Systems Employing Additional Electrodes.

a small spot. However, the spot size depends upon the grid bias, which is very undesirable.

An accelerating grid, or other element, in the first lens is often used to make the beam current independent of the second-lens focus. Two forms of multi-grid first-lens systems are shown in Fig. 13.14.

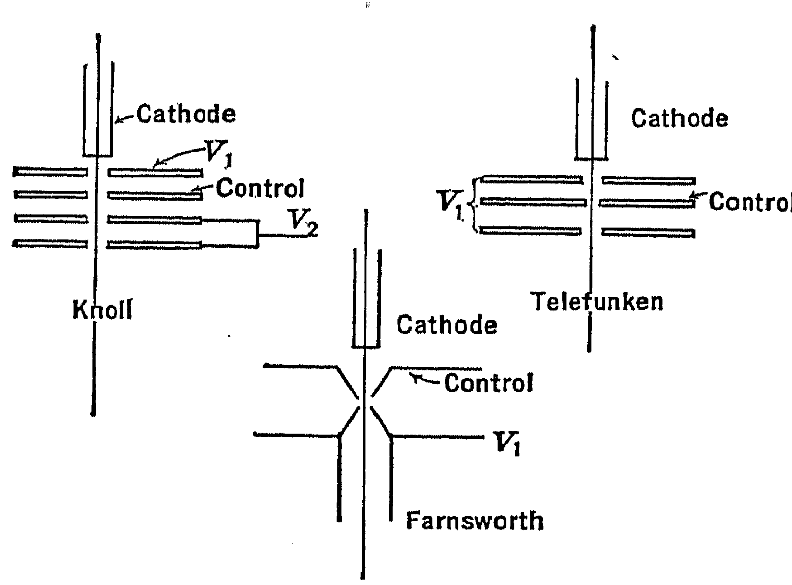


Fig. 13.15.—First-Lens Systems with Crossover Apertures.

An aperture at the crossover, as was pointed out in the preceding section, is advantageous, in that it gives a sharply defined object which is imaged by the second lens onto the screen. Several forms of first-lens systems constructed with crossover apertures are shown in Fig. 13.15.

A very effective first lens system, designed by Law, is illustrated in Fig. 13.26. This system makes use of a curved cathode to minimize aberration in the first lens, as well as an aperture at the crossover. Several additional apertures are provided between the crossover and cathode so that the position of the crossover can be accurately controlled.

13.6. The Second Lens. The second lens system images the crossover into the spot on the viewing screen, or mosaic. In general, this lens is relatively simple, and the first-order-image properties can be readily calculated by the methods given in Chapter 4.

A number of alternative forms are possible for an electrostatically focusing second lens. Some of the principal configurations are illustrated

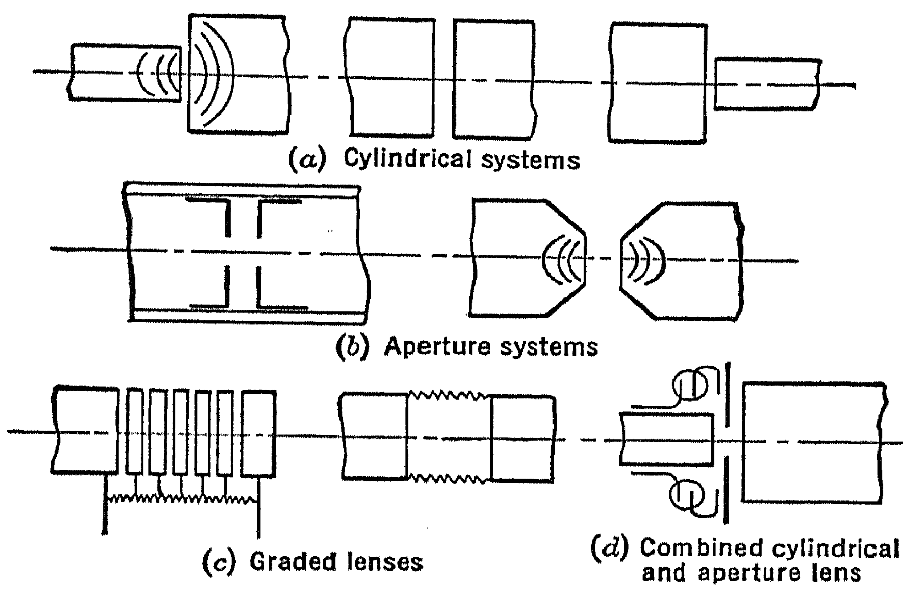


FIG. 13.16.—Various Second-Lens Configurations.

in Fig. 13.16. By a proper choice of potentials and diameters, any of these lenses can be made to have suitable first-order properties, and many of them have been very successful in practical guns.

The two major factors governing the selection of the most suitable lens from among the possible forms are spherical aberration and ease and accuracy of construction. The first group (a) of Fig. 13.16 has been investigated experimentally by Epstein* and Gundert† to determine spherical aberration. Of the two diameter ratios studied by the first-named investigator, a diameter ratio of 1.5 was found to give more aberration than a ratio of 3.6. In each case the diameter of the second anode was the greater. Gundert carried the investigation further and found that spherical aberration could be made to go through a minimum by making the first anode of greater diameter than the second. This cor-

* See Epstein, reference 10.

† See Gundert, reference 14.

rection is possible because the optimum ratio balances the defect in the converging portion of the lens against the defect in the diverging portion. Cylindrical lens structures of this type can be simply and easily constructed.

The second group has similar properties if the two apertures have different sizes. In this group the second configuration shown is of interest in that the equipotential surfaces are nearly spherical. The aberration in this lens is very severe, much more so than for two flat apertures, or cylinders, of the same diameter.

The last lens shown combines an aperture second anode with a cylindrical first anode. A proper selection of diameters permits a partial correction of spherical aberration. This lens is very easy to assemble and align because the second anode is rigidly attached to the first anode in such a way that the whole structure can be mounted on a gun press.

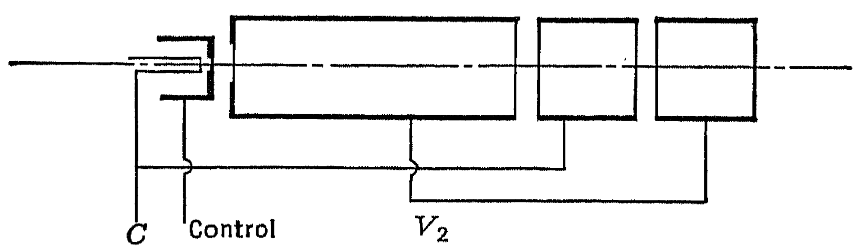


Fig. 13.17.—Gun with Unipotential Second Lens.

The spherical aberration for a given diameter of the beam through the second lens can be reduced by increasing the size of the lens. In electrostatic focusing, the physical size of the lens is limited by the diameter of the gun neck which, in turn, is determined by the deflecting system employed (assuming magnetic deflection). With a magnetic second lens no such limit of size exists. Other things being equal, it is therefore possible to obtain a lens having less spherical aberration with magnetic focusing. A large lens has, it is true, certain practical disadvantages, such as the length of tube necessary to separate the deflecting and the focusing fields. However, a number of workers have obtained very satisfactory results with this type of lens. The afore-cited Law projection tube has such a lens, which is illustrated in Fig. 13.27.

There are obviously many other possible second lens configurations with either electrostatic, magnetic, or combined fields. Some of these are as yet untried; others have been tested with varying degrees of success. An interesting case is that in which the second focusing field is a unit lens (Einzellinse) arranged in such a way that only a single voltage need be applied to the gun. Of the several ways of realizing this condition one is illustrated in Fig. 13.17. A system of this type has been found by Iams

to function very satisfactorily. The advantage of such a gun lies in the fact that it simplifies the power supply and that it is not defocused by voltage changes. On the other hand, it requires much more careful gun assembly, is more difficult to correct for defects in the second lens, and lacks flexibility.

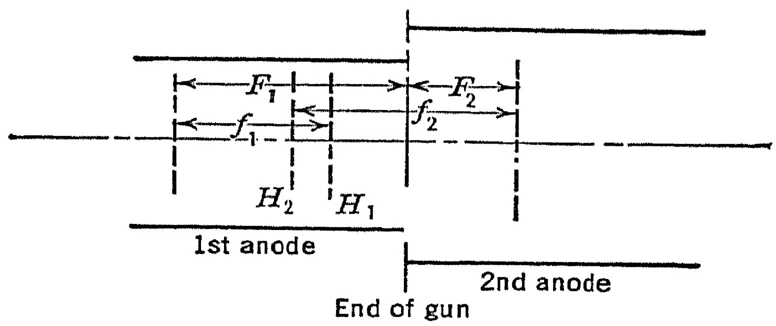


Fig. 13.18.—Cardinal Points of Two-Cylinder Lens.

Returning to the more conventional gun, the design procedure for the second lens has been very ably described by Epstein.* The cylindrical second focusing systems illustrated in group (*a*) of Fig. 13.16 form the basis of his work. The cardinal points of a number of cylindrical lenses having different diameter ratios were calculated for a wide range of voltage ratios from potential maps made with a plotting tank. These determinations were then verified in experimental guns. The characteristics of these second focusing fields were those of thick lenses having crossed principal planes. The arrangement of the four cardinal points is shown in Fig. 13.18. Fig. 13.19 reproduces two sets of curves, giving the focal lengths and positions of the two focal points for systems having diameter ratios of 1.5 and 0.44 over a range of voltage ratios from 3.5 to 6.5.

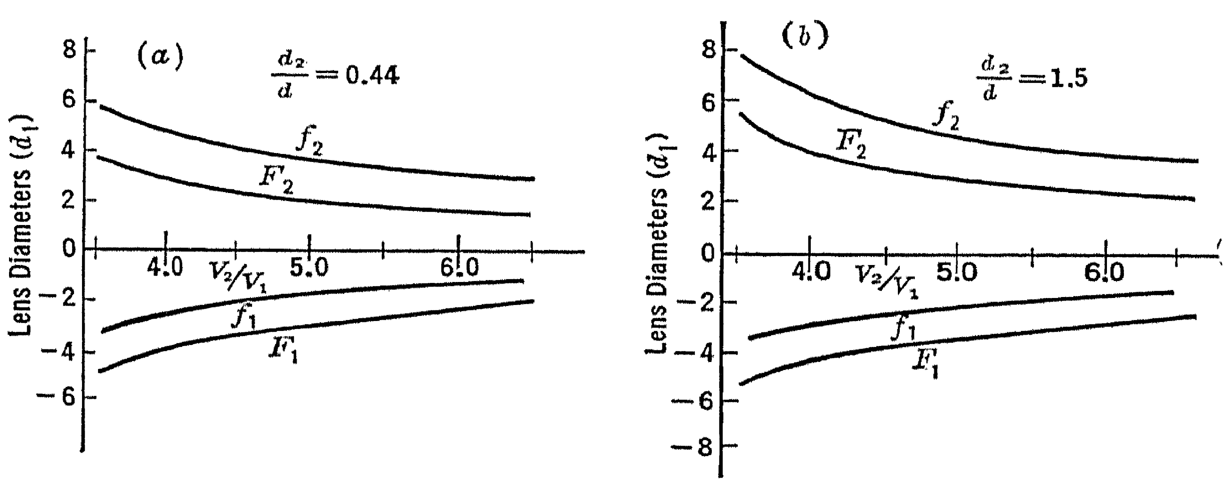


Fig. 13.19.—Optical Constants of Two-Cylinder Second Lens (Epstein).

* See Epstein, reference 10.

Lengths given in these curves are measured in terms of gun diameters, that is, in terms of the diameter of the first anode, and the focal points are located relative to the end of the first anode (often spoken of as the "end of the gun"). Similar curves for the diameter ratio unity are given in Fig. 4.15. With the aid of these curves and the ordinary "thick lens" formulas, it is possible to locate the end of the gun from a knowledge of the distance between the crossover and the screen, and the focusing voltage ratio. When a simple first lens system is used, such as is shown in Fig. 13.9, the crossover can be assumed to be at the cathode, thus simplifying the design problem. By a further application of the "thick lens" relations, the magnification of the spot (i.e., ratio of spot diameter to crossover diameter) can be determined.

13.7. The Iconoscope Gun. As has already been stated, the principal requirement of the Iconoscope gun is a very small, sharply defined spot. The current required in the beam is very small, being of the order of 0.1 or 0.2 microampere, and furthermore the control characteristics of the grid are relatively unimportant. The main consideration, therefore, is spot size, even at the expense of current and control.

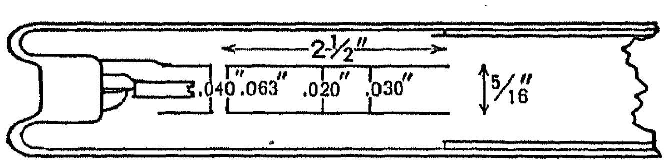


FIG. 13.20.—Simple Iconoscope Gun.

The first lens system can satisfactorily be of the simple type shown in Fig. 13.9; however, a more complicated system, such as one employing an accelerating grid, is often found convenient as this permits adjusting the beam current without defocusing the spot. The grid and first anode apertures must be small, and usually the cathode area is very much restricted.

A simple cylindrical second lens system, such as is illustrated in group (*a*) of Fig. 13.16, can be used. Because of the small current required, the aperture of the second lens can be made very small, thus minimizing spherical aberration.

If a single aperture is used in the first anode cylinder to reduce the beam diameter, secondary electrons generated at the rim of the aperture will be drawn through the second lens by the second anode and will reach the mosaic. This will give a foggy appearance to the reproduced picture. To avoid this difficulty, one or more apertures which are slightly larger than the beam diameter are placed between the limiting aperture and the second lens.

A complete gun which gives satisfactory performance in an Iconoscope is shown in Fig. 13.20. The cylindrical metal parts are made of seamless nickel tubing. Tantalum, nichrome, or nickel bronze are often preferred

because of their non-magnetic properties. Lavite, or other suitable insulating material, is used to separate the grid from the first anode in such a way that these two elements are rigidly and accurately spaced and aligned. The whole assembly is mounted on the gun press which is then sealed into the tube.

A photograph of the gun assembly for the normal Iconoscope is shown in Fig. 13.21. This gun is slightly more complicated than that just described, having a combined cylinder and aperture second lens, an accelerating grid first focusing system, etc.

Because of its small beam current, the performance of the Iconoscope gun with reference to spot size cannot be conveniently measured on a fluorescent screen. In order to make this measurement, the gun is mounted in a regular Iconoscope blank which contains instead of a mosaic a specially prepared screen. This screen consists of a sheet of mica,

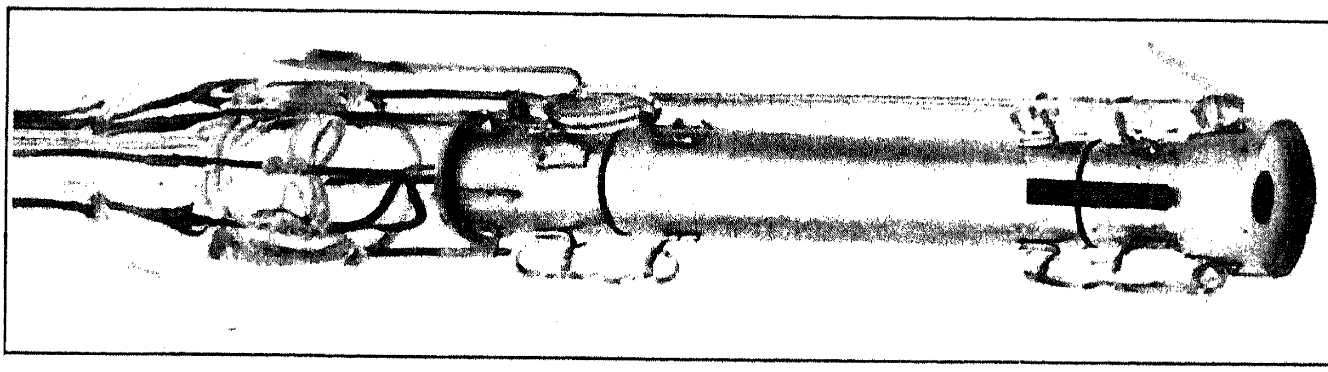


FIG. 13.21.—Photograph of a Typical Iconoscope Gun.

or glass, coated with a metallic film through which have been ruled groups of fine vertical lines. Each group has a different line separation; for example, 0.002, 0.003, 0.004, and 0.005 inch. When this screen is scanned and the signal output observed on the screen of a Kinescope, the line pattern will be reproduced if it is resolved by the spot. By decreasing the horizontal amplitude of scanning in the test tube, the pattern of the lines can be magnified on the Kinescope screen, thus eliminating the possibility of the resolution of the reproduced pattern being limited by the viewing tube or amplifier system, rather than the gun under test.

The performance of the gun can be estimated by observing which groups of lines are resolved under these conditions. A correct estimate of the resolution by this method requires some experience, and it is quite common for a beginner to assign too small a value to the spot diameter. This is due to the fact that lines which are only partially resolved because of the non-uniform current distribution in the spot can easily be mistaken for truly resolved lines.

13.8. The Kinescope Gun. While the greater current demand and the need for good control characteristics increase the difficulty of the design problem of the Kinescope gun, they are partially offset by the larger permissible spot size and the higher overall voltage.

A simple gun, similar to that shown in Fig. 13.20, will function admirably in the Kinescope if large apertures and some alteration in the spacings of the first lens are made to permit the required higher beam current. More complicated guns, however, are advantageous in that they can be made to give less defocusing of the spot with grid swing, thus permitting more complete utilization of the current emitted by the cathode.

The size of the spot required is determined by the size of the viewing screen, the number of scanning lines, and the overlap desired. For example, a 12-inch Kinescope will produce a picture which is about $7\frac{1}{2}$ by 10 inches in size. Therefore, the separation of lines in a 441-line pattern will be 0.016 inch, or about 0.4 mm. In order that two lines do not run together, the brightness at a distance of 0.2 mm from the center of a line must be less than a certain predetermined fraction of that at the center of the line. For the present illustration, let it be assumed that this ratio is 0.01. From these data the "spot diameter" can be determined.

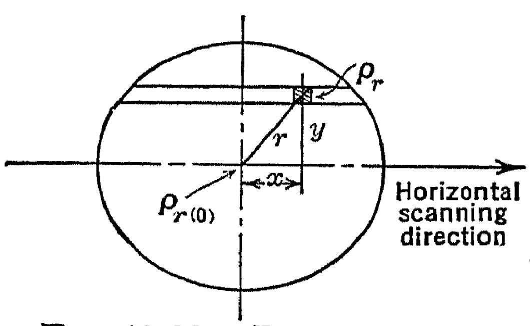


FIG. 13.22.—Determination of Current Distribution in Scanning Line.

As has already been pointed out, the term "spot diameter" is rather meaningless unless arbitrarily defined. In section 13.3, the diameter was defined as that within which the ratio of current density to current density at the center was greater than a certain fraction, K. It is apparent that the diameter of the spot so defined will not necessarily equal the line width, because the line is being scanned with a circular spot. In order to determine the diameter, it is necessary to integrate the density in the spot over elemental strips parallel to the scanning line as indicated in Fig. 13.22. If the current distribution in the spot is given by Eq. 13.11, derived for the ideal gun, the density can be written in the following form:

$$\rho = A\epsilon^{-Br^2}.$$

From the figure:

$$r^2 = x^2 + y^2.$$

Therefore the radial density becomes:

$$\rho = A\epsilon^{-B(x^2+y^2)},$$

and the average current density $\rho_L(y)$ in an elementary strip parallel to the line at a distance y from the center of the spot is:

$$\rho_L(y) = 2\int_0^\infty A\epsilon^{-By^2}\epsilon^{-Bx^2}dx = A\sqrt{\frac{\pi}{B}}\epsilon^{-By^2}.$$

The ratio of current density at a distance y from the center of the line to that at the center $\rho_L(0)$ will, of course, be:

$$\frac{\rho_L(y)}{\rho_L(0)} = \frac{A\sqrt{\frac{\pi}{B}}}{A\sqrt{\frac{\pi}{B}}}\epsilon^{-By^2} = \epsilon^{-By^2}.$$

If the "line width" 2δ is defined by the relation

$$\frac{\rho_L(\delta)}{\rho_L(0)} = K,$$

a comparison with Eq. 13.14 shows that it is equal to the "spot diameter" 2β. Therefore the spot diameter should be 0.4 mm for a 12-inch screen if no overlap is desired, or 0.8 mm for a 50 per cent overlap.

The current required in the spot depends upon the type of fluorescent material used on the screen, the bombarding voltage, and the brightness required. With white zinc sulphide as the screen material and 6000 volts on the second anode of the gun, 200 microamperes will produce sufficient brightness so that the picture on a 12-inch screen can be easily viewed in a normally lighted room.

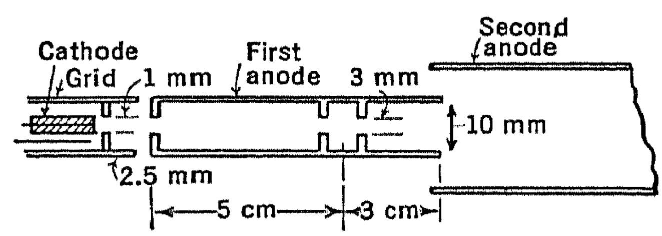


Fig. 13.23.—Elementary Kinescope Gun.

The first lens must be designed to give the required current and also to have suitable control characteristics. In the simple Kinescope gun illustrated in Fig. 13.23 (note its similarity to that shown in Fig. 13.20), the size of the grid aperture is important as it has a great influence on the spot size. Its diameter should, in the gun under consideration, be of the order of 1 mm. While the grid skirt has relatively little influence on the spot size, it has a large effect on the beam current and control characteristics. In the gun shown, a skirt length of 2.5 mm will give a cutoff at about 50 volts under normal operating conditions.

The aperture stop in the second focusing field must be much larger than that used in the Iconoscope gun. However, in order to reduce the

spherical aberration of the second lens to the same order as the other aberrations in the gun, it should restrict the aperture of the second lens to 25 per cent of the gun diameter. Since the unrestricted beam would occupy about 40 per cent of the gun diameter at the center of the lens, a current equal to about three halves the beam current will be absorbed by the first anode. This, of course, is wasted power and merely increases the load on the first anode voltage supply in addition to heating the first anode. As in the Iconoscope, more than one stop is necessary in order to prevent scattered electrons from entering the beam.

Although the gun described gives a fairly satisfactory performance, it has been superseded by more complicated forms. One of these is shown in Fig. 13.24. This gun is similar in principle to that described but has a somewhat different form of the first and second lens which makes its

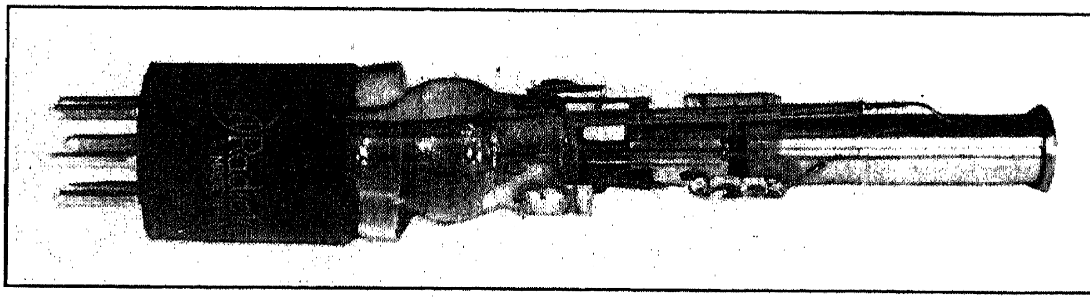


Fig. 13.24.—Photograph of a Typical Kinescope Gun.

performance better from the standpoint of power absorbed by the first anode, defocusing of the spot by the control voltage, and the form of the control characteristics.

13.9. The Projection Gun. The design of the projection tube gun presents many more difficulties than that of any of the guns heretofore described. The beam current required is greater than that of a direct-viewing Kinescope, its spot size must be nearly as small as that of the Iconoscope gun, and it must have reasonable control characteristics. That the overall voltage is usually many times as great as that for the direct-viewing tube is an advantage as far as obtaining a small spot is concerned but introduces additional insulation difficulties.

As was pointed out in the preceding chapter, it is convenient optically to use a small screen. For example, a screen size of $2\frac{1}{4}$ by 3 inches was cited as being found satisfactory. The separation between lines in this case is about 0.1 mm. Therefore, for a 50 per cent overlap, the spot size should be about 0.2 mm, the diameter being defined as before. At 15,000 volts applied to the gun the current required is 0.5 milliampere or more.

A developmental projection gun based upon the simple design illustrated in Fig. 13.23, with a three-element first lens and a two-element cylindrical second focusing system, has been found to function fairly satisfactorily. Insulation precautions, of course, are necessary to avoid cold emission and electrical leakage, and the edges of all lenses across which a high voltage appears are rolled. The control characteristics and spot size, when the gun is operated at 10,000 volts, are shown in Fig. 13.25. When this gun is biased so as to deliver 0.4 milliampere into the spot, the current lost to the first anode is about 1.0 milliampere.

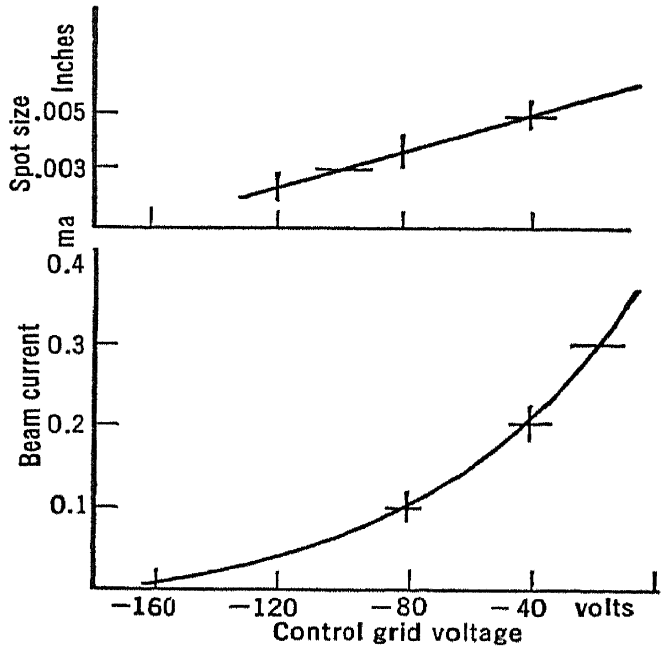


Fig. 13.25.—Control Characteristics and Spot Size of Simple Projection Gun.

By altering the first focusing field to permit a greater effective cathode area and the use of a multi-grid system, and also by having a larger-diameter second lens, more efficient electrostatic projection guns have been successfully constructed.

Mention has already been made of a very successful projection gun designed by Law, having a magnetic second lens and an aperture at the crossover of the first lens. This gun* is designed to operate at 10,000 to 20,000 volts and will deliver a 2-milliampere beam current into a spot 0.3 mm in diameter. The size of the screen used with this gun is 1.8 by 2.4 inches.

Details of the first lens system are shown in Fig. 13.26. The cathode is curved rather than flat, and has an oxide-coated emitting area about 0.05 inch in diameter. Four apertured disks are located in front of the cathode; the fourth, which has an opening only 0.004 inch in diameter, is the limiting aperture for the crossover. The first two are electrodes for modulating the beam current. Electrode 3 is placed at such a potential and position as to bring the crossover exactly at the limiting aperture. The electrons have acquired their full velocity by the time they pass through the crossover.

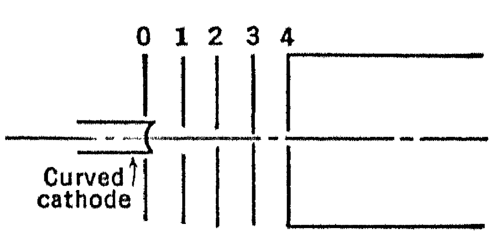


Fig. 13.26.—First Lens of Projection Gun (Law).

The complete gun is shown in Fig. 13.27. The second lens is a specially

* See Law, reference 12.

designed electromagnet, so shaped as to reduce spherical aberrations. A magnetic lens is used because larger aberration-free apertures can thus be obtained without sacrificing ease of deflection when a magnetic de-

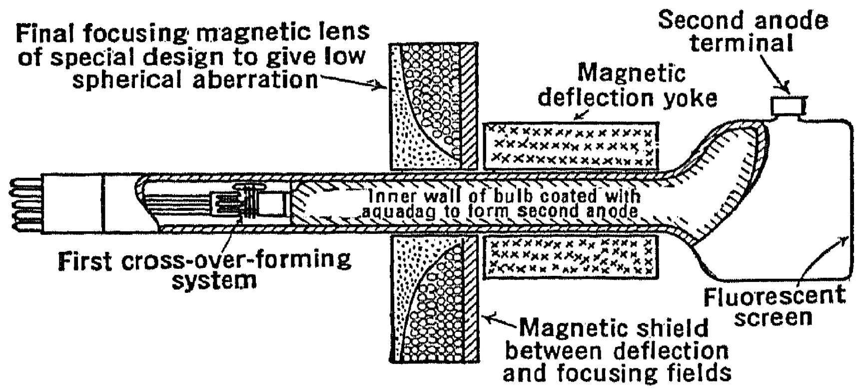


Fig. 13.27.—Diagram of Assembly of Law's Projection Gun.

flecting system is used. The lens shown in the figure exhibits negligible spherical aberration over a 6-mm aperture.

The control characteristics of this gun are shown in Fig. 13.28.

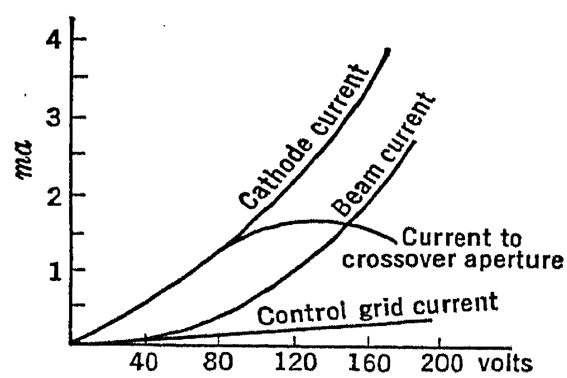


Fig. 13.28.—Current Characteristics of Law's Projection Gun.

13.10. Defects of the Electron Gun. The principal defects of the electron gun come under two general headings:

1. Structural defects.
2. Aberrations in the electron-optical systems.

Defects of the first class are usually due to misalignment of the gun parts and to apertures and cylinders which are out of round. These faults manifest themselves by giving a spot which is elliptical rather than circular, or by giving an enlarged, poorly defined spot. The remedy, of course, is greater care in the construction of the gun.

The aberrations of the electron-optical systems have been discussed in brief in the description of the various gun elements. Chief among these aberrations are:

Chromatic aberration in the first lens.
Space-charge defect in the first lens.
Spherical aberration in the first and second lenses.

These aberrations can be corrected, in part, by the proper design of the various gun elements and suitably disposed limiting apertures. However, it is evident from the foregoing that, although guns meeting the present television requirements can be made, nevertheless gun design is far from a closed field. Much has yet to be learned about the behavior of the electrons in the first focusing field, especially with regard to the effects of space charge, and about the nature and correction of the spherical aberration in the second lens.

Only through a more critical and complete study of electron optics will this information become available. In view of the rapidly increasing interest in this field, optimism is justifiable.

REFERENCES

1. E. Brüche and O. Scherzer, "Elektronenoptik," Julius Springer, Berlin, 1934.
2. I. G. Maloff and D. W. Epstein, "Electron Optics in Television," McGraw-Hill, New York, 1938.
3. F. Schröter, "Fernsehen," Julius Springer, Berlin, 1937.
4. A. Wehnelt, "Increase in the Sensitiveness of the Braun Tube by Use of Cathode Rays of Lower Velocity," *Physik. Z.*, Vol. 6, p. 732, 1905.
5. E. Brüche and O. Scherzer, "Braun Tube as an Electron Optical Problem," *Z. tech. Physik.*, Vol. 14, pp. 464–466, 1933.
6. M. Knoll, "Electric Electron Lens for Cathode Ray Tubes," *Arch. Elektrotech.*, Vol. 28, pp. 1–8, January, 1934.
7. M. Knoll and J. Schloemilch, "Electron Optical Current Distribution in Electron Tubes with Control Electrodes," *Arch. Elektrotech.*, Vol. 28, pp. 507–516, August, 1934.
8. I. G. Maloff and D. W. Epstein, "Theory of Electron Gun," *Proc. I. R. E.*, Vol. 22, pp. 1386–1411, December, 1934.
9. M. Knoll, "Electron Optics in Television Technique," *Z. tech. Physik.*, Vol. 17, pp. 604–617, 1936.
10. D. W. Epstein, "Electron Optical System of Two Cylinders as Applied to Cathode Ray Tubes," *Proc. I. R. E.*, Vol. 24, pp. 1095–1139, August, 1936.
11. D. B. Langmuir, "Limitations of Cathode Ray Tubes," *Proc. I. R. E.*, Vol. 25, pp. 977–991, August, 1937.
12. R. R. Law, "High Current Electron Gun for Projection Kinescopes," *Proc. I. R. E.*, Vol. 25, pp. 954–976, August, 1937.
13. V. K. Zworykin and W. H. Painter, "Development of the Projection Kinescope," *Proc. I. R. E.*, Vol. 25, pp. 937–953, August, 1937.
14. E. Gundert, "Electron Lens Aberrations Shown by Thread Beams," *Physik. Z.* Vol. 38, pp. 462–467, June, 1937.

CHAPTER 14

VIDEO AMPLIFIERS

14.1. The Amplifier. In the foregoing chapters the two basic terminal tubes of the television system have been discussed in some detail. It is evident from this discussion that, even where the Iconoscope merely operates a monitoring Kinescope, it is necessary to amplify its signal output by a factor of many thousands before the signal voltage will be sufficient to swing the control grid of the viewing tube. If the Iconoscope is an element in a complete television broadcast system, even further amplification will be necessary. Not only must the picture signal from

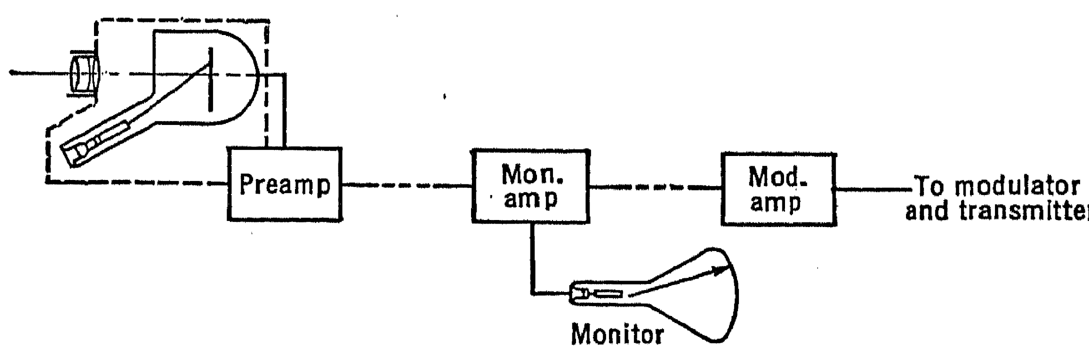


Fig. 14.1.—Video Amplifier Chain at Transmitter.

the Iconoscope be amplified up to the modulation level of the transmitter, but also the signal must be amplified at the receiver to the level required to operate the Kinescope.

The amplification of the transmitter may, for convenience, be divided into three steps, as was mentioned in Chapter 7. A pre-amplifier with a gain of about 1000 is located at the Iconoscope camera. The final stage of this amplifier has a fairly low output impedance and feeds a low-capacity flexible cable. The input of the monitoring amplifier is fed from this cable and the signal is amplified up to monitoring level, and also to a level sufficient to supply the signal to a coaxial cable connecting the monitor to the transmitter. The output of the second cable supplies the modulation amplifier which amplifies the signal to such an extent that it can be used to modulate the transmitter. This chain is illustrated in Fig. 14.1.

At the radio receiver the incoming signal must again be amplified.

Here the problem is somewhat different in that a large part of the amplification takes place in the radio-frequency and intermediate stages of a superheterodyne receiver. This is shown in Fig. 14.2. However, the output of the second detector must usually be amplified by a video amplifier, similar to those at the Iconoscope, before the signal is sufficient to operate the Kinescope control grid.

Since the requirements of these video amplifiers are the same except for the power they must handle, the principles which will be considered in this chapter apply equally to all of them.

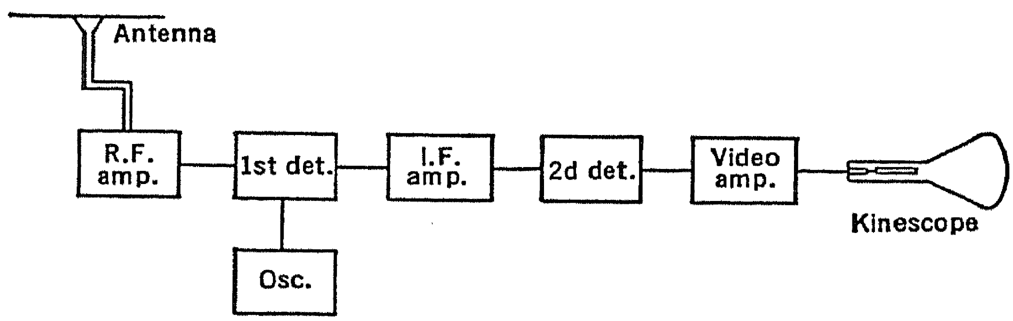


FIG. 14.2.—Amplifiers of Superheterodyne Receiver.

14.2. Requirements of a Video Amplifier. The video amplifier is a thermionic vacuum-tube amplifier, and its basic principles are similar to those of the ordinary audio amplifier familiar in radio practice. Because the general amplifier problem, as represented by the audio amplifier, has been treated so ably and completely in numerous textbooks, it is unnecessary to discuss the entire theory of the amplifier in the following consideration. Instead, only problems resulting from differences in the requirements of audio and video amplifiers will be treated.

The two principal demands which distinguish the video amplifier from the audio amplifier are:

1. A constant frequency response over a band several megacycles wide.
2. A constant time delay over the usable frequency range.

The differences between the two types of amplifiers arise from these requirements of the video amplifier. Some attendant problems which must be considered are those of maintaining response at both high and low frequencies and at the same time keeping a linear variation of phase delay with frequency, of shielding without introducing high capacity, of decoupling the power supply to avoid low-frequency oscillation or motor boating, and of keeping a satisfactory signal-to-noise ratio.

The frequency band required to transmit a picture was discussed in

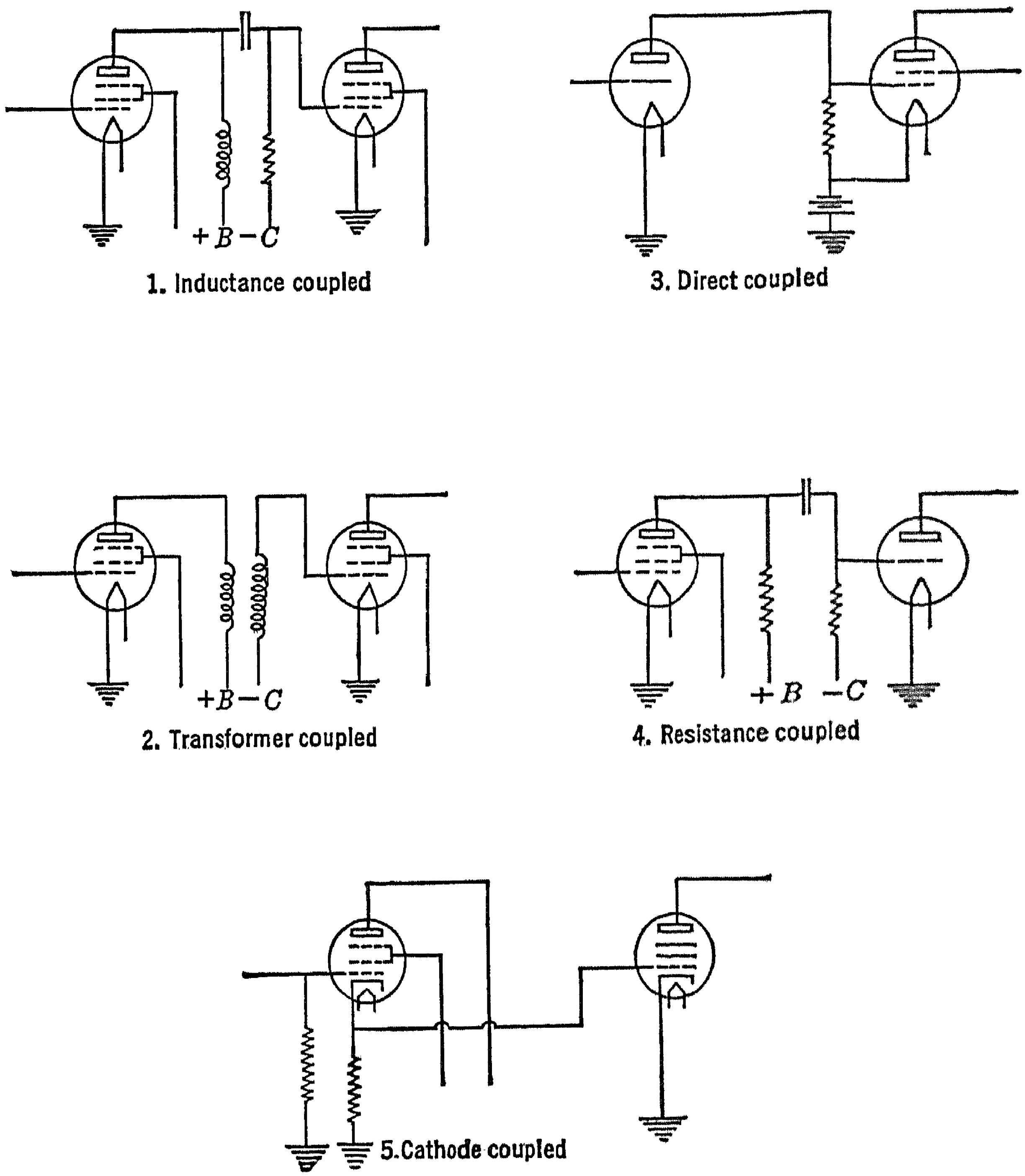


FIG. 14.3.—Principal Types of Class A Amplifier Couplings.

some detail in Chapter 6. Here it was concluded that to transmit a 441-line picture with an aspect ratio of 4×3 at 30 frames per second required a bandwidth of no less than 3 megacycles and preferably 4.5 megacycles. It should be noted that most audio amplifiers have a constant response over less than 10,000 cycles or, in other words, have a bandwidth of less than one three-hundredth of that required of a video amplifier.

Since it is a property of the human ear that it cannot (within rather wide limits) detect phase distortion, the problem of maintaining constant time delay does not enter into the design of an audio amplifier. However, it is apparent from the Fourier analysis of the picture signal given in Chapter 6 that the phase of the various frequency components making up the picture is very important. This is also obvious if the mechanism of the formation of a picture by scanning is kept in mind. The scanning beam in a 12-inch Kinescope moves across the screen at a speed of 1.3×10^5 inches per second. Thus, if there is a time delay of only 1 microsecond for a particular frequency, the position of light maxima and minima resulting from this frequency will be shifted by 0.13 inch. If there is an equal time delay for every frequency it will merely result in a shift of the picture as a whole and does not introduce any distortion. Therefore, the time delay over the frequency band used must be zero or a constant. Time delay is related to the phase angle as follows:

$$\Delta t = \frac{\phi}{2\pi f},$$

where ϕ is the phase angle. Hence a constant time delay requires a linearly increasing or decreasing phase angle. The phase angle of the video amplifier must therefore be a linear function of the frequency over the 3- to 4.5-megacycle band required.

14.3. Amplifier Types. A rather wide variety of types of amplifying systems for sound equipment is available, differing from one another in the means used to couple the successive vacuum tubes. The principal types are:

1. Inductance coupled.
2. Transformer coupled.
3. Direct coupled.
4. Resistance coupled.
5. Cathode coupled.

This classification is not truly justified, as will become apparent as the discussion proceeds; however, it simplifies a preliminary survey of the

field. The five amplifier types are shown schematically in Fig. 14.3. There are, of course, many modifications of each.

Upon examining each in turn as to its desirability as a video amplifier, it becomes evident that the resistance-coupled amplifier most nearly meets the requirements.

The inductance-coupled amplifier, though economical in power consumption, is impractical because the great range of frequencies involved makes it impossible to use any of the known types of coupling chokes. Air-core chokes cannot be used because the very large number of turns necessary to obtain sufficient impedance at low frequencies makes the distributed capacities high, which spoils both phase and amplitude response at high frequencies. Iron-core chokes, on the other hand, exhibit a lack of response at high frequencies and therefore are not suitable. Magnetite cores, or those using a powdered iron alloy in an insulating binder, come closer to yielding chokes which have the desired properties, but even these have undesirable phase shifts and variation in impedance over a band as wide as that required by television transmission.

Transformer-coupled amplifiers suffer from the same type of defects found in inductance-coupled amplifiers and are therefore also impractical.

Direct-coupled amplifiers are those in which each tube as a whole is maintained at a higher potential than the preceding tube, so that the plate of one tube is at the d-c potential of the grid potential of the next. This type of amplifier can be made to have excellent phase and amplitude characteristics. However, its adjustment is extremely critical, and, furthermore, it is very sensitive to fluctuations of the voltage supplying it. Even though this type of amplifier has been successful in experimental installations, its operation is considered too unreliable to make it really serviceable for a practical television system.

The resistance-coupled amplifier is by no means ideal; nevertheless by resorting to certain corrective measures it is possible to obtain a response which is satisfactory. This type of amplifier is that chiefly used as video amplifier and, therefore, will be considered in detail in the succeeding sections.

The final class, namely, the cathode-coupled amplifiers, cannot be used for voltage amplification since their gain never exceeds unity. However, they have their place in television as a means of coupling a low impedance load, such as a cable, to a voltage amplifier.

14.4. Resistance-Coupled Amplifier. The principle of the operation of the resistance-coupled amplifier can be seen from the circuit diagram in Fig. 14.3. If a positive impulse is impressed upon the grid of the first tube through its coupling condenser, the plate current through this tube

increases. The rise in current increases the IR drop through the external plate resistance of the tube, causing its plate end to become more negative. The negative impulse thus produced is transmitted through the second coupling condenser onto the grid of the second tube, where it causes a decrease in the plate current of that tube.

In order to predict the behavior of a resistance-coupled amplifier as a video amplifier, a quantitative analysis is necessary. A single stage of the amplifier, together with its equivalent circuit, is shown in Fig. 14.4. The symbols in this figure have the following significance. The input voltage is e_1, and the voltage amplification μ, so that in the equivalent circuit the first tube can be represented as a voltage generator whose output is $e_1\mu$. R, r_p, and R_g are the external plate resistance, the internal plate resistance, and the grid resistance, respectively. The condenser C is the

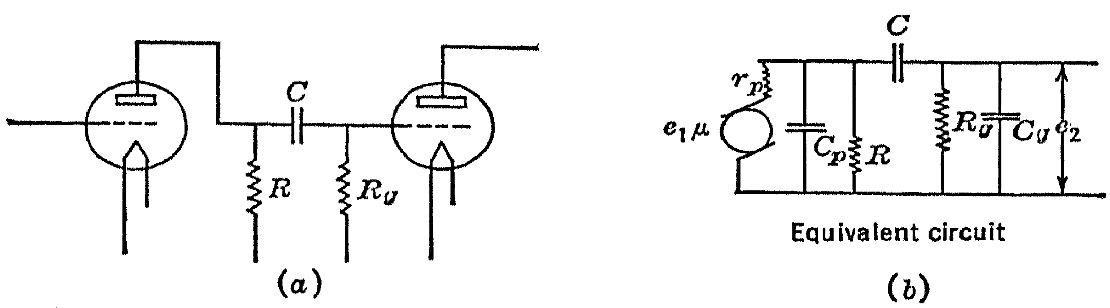


Fig. 14.4.—Simple Resistance-Coupled Stage and Its Equivalent Circuit.

coupling condenser, C_p is the capacity of the plate of the first tube to ground, and C_g is the total input capacity of the succeeding tube.

It is evident from Fig. 14.4 that the interstage coupling in a resistance-coupled amplifier is an impedance network rather than a pure resistance. Similarly, for the other amplifiers shown in Fig. 14.3, the coupling is a network and the designated coupling is merely a convenience, indicating the predominant character of the network over part or all of the working range of the amplifier. The coupling network of the resistance-coupled amplifier exhibits the characteristics of a resistance only over the relatively small fraction of the total frequency band for which the impedance of the condenser C is small compared with R and the reactance of the tube capacity is large in comparison with R and r_p.

The response of this network to a voltage applied to the grid of the first tube can be determined from the circuit constants. If a voltage e_1 is applied at the first grid, the voltage e_2 appearing at the grid of the second tube will be:*

$$e_2 = \frac{Z_1Z_2\mu e_1}{Z_1Z_2 + r_pZ_1 + r_pZ_2 - j(r_pX_c + X_cZ_2)}, \tag{14.1}$$

* To conform with electrical engineering practice, j is used instead of i to express $\sqrt{-1}$ in this and in subsequent chapters.

where

$$Z_1 = \frac{R_g - j\omega C_g R_g^2}{1 + \omega^2 C_g^2 R_g^2},$$

$$Z_2 = \frac{R - j\omega C_p R^2}{1 + \omega^2 C_p^2 R^2},$$

and

$$X_c = \frac{1}{\omega C}.$$

It is apparent that the output voltage e_2 is not independent of the frequency $f = 2\pi\omega$, and that therefore the response of the amplifier may not be flat over the frequency band required by the video signal.

Since for an amplifier of this type $C \gg C_p$ or C_g, and R_g is very large, the expression given in Eq. 14.1 reduces to:

$$e_2 = \frac{\mu e_1 R}{R + r_p}$$

over the middle range of frequencies (i.e., frequencies such that $1/\omega C_g$ and $1/\omega C_p$ are large compared with R, and $1/\omega C$ is small in comparison with R). At the two extremes e_2 falls off for a given input voltage and, furthermore, there is a varying phase relation between the input and output voltages.

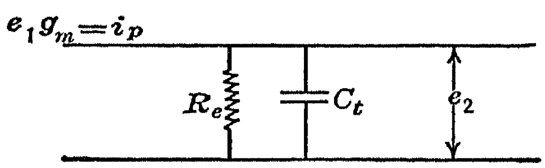


Fig. 14.5. — High-Frequency Equivalent Circuit of a Resistance-Coupled Stage.

Instead of evaluating Eq. 14.1 directly, which leads to a rather cumbersome algebraic expression, more information as to the effect of the various circuit constants on the frequency response is obtained if approximations are made which permit a simple evaluation of the response at the highest and lowest frequencies within the transmitted band.

It is well known that an amplifier tube may be represented by a constant current generator shunted by the internal impedance of the tube. Representing the transconductance by g_m, the current output of this generator is given by the relation

$$i_p = e_1 g_m$$

Furthermore, since the television amplifier must operate at low frequencies, the coupling condenser C must be large and will, therefore, present a negligible impedance to the upper frequencies. At these frequencies the equivalent circuit can be represented with fair accuracy by

Fig. 14.5. In this figure $C_t = C_p + C_g$ is the total shunt capacity, and R_e is the effective shunt resistance of R, r_p, and R_g in parallel, namely:

$$\frac{Rr_pR_g}{Rr_p + R_gr_p + R_gR}.$$

The output voltage e_2 under these conditions is

$$\begin{aligned} e_2 &= e_1g_mZ \\ &= \frac{R_e - j\omega C_tR_e^2}{1 + \omega^2C_t^2R_e^2}e_1g_m. \end{aligned} \tag{14.2}$$

The magnitude of the gain as a function of frequency is, therefore:

$$G = \frac{R_eg_m}{\sqrt{1 + 4\pi^2f^2C_t^2R_e^2}}, \tag{14.3}$$

while the phase angle is:

$$\phi = -\tan^{-1}2\pi fC_tR_e. \tag{14.4}$$

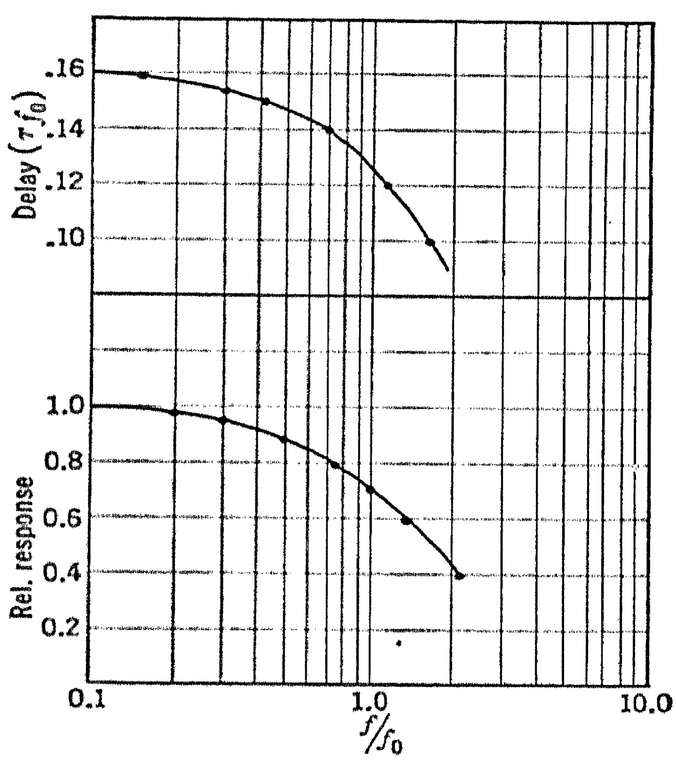


FIG. 14.6.—Response of the Coupling Shown in Fig. 14.5.

For convenience, an arbitrary maximum or critical frequency f_0 defined by:

$$f_0 = \frac{1}{2\pi C_tR_e} \tag{14.5}$$

will be assumed. Later, when considering the correction of the amplifier

at high frequencies, the reason for this will become apparent. The magnitude and phase angle of the response in terms of this critical frequency are:

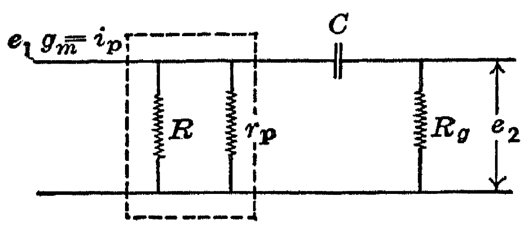


FIG. 14.7. — Low-Frequency Equivalent Circuit of a Resistance-Coupled Stage.

$$G = \frac{R_e g_m}{\sqrt{1 + (f/f_0)^2}}, \tag{14.6}$$

$$\phi = -\tan^{-1}\frac{f}{f_0}. \tag{14.6a}$$

These are shown plotted as a function of frequency in Fig. 14.6.

At very low frequencies the impedance of the coupling condenser C cannot be considered negligible, while the shunt load impedance due to the capacity of the tube and wiring becomes so high that it can be neglected. The equivalent circuit for this case

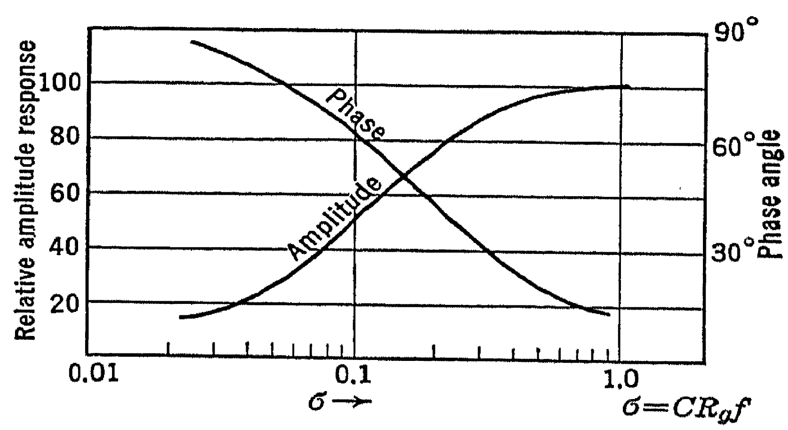


FIG. 14.8.—Response of the Coupling Shown in Fig. 14.7.

will, therefore, be that given in Fig. 14.7. The output voltage of this circuit will be:

$$e_2 = e_1 g_m R_a \frac{R_g}{(R_g + R_a) - jX_c},$$

where

$$R_a = \frac{Rr_p}{R + r_p}.$$

Since R_a, in this type of amplifier, is small compared with R_g,

$$e_2 = e_1 g_m R_g R_a \left(\frac{R_g + jX_c}{R_g^2 + X_c^2}\right). \tag{14.7}$$

The magnitude of the gain is, therefore:

$$G = g_m R_g R_a \frac{1}{\sqrt{R_g^2 + X_c^2}}, \tag{14.8}$$

and the phase angle:

$$\phi = \tan^{-1}\frac{X_c}{R_g}, \tag{14.9}$$

where $X_c = 1/2\pi fC$. These two functions are shown in Fig. 14.8.

It is clear from the foregoing that it may be necessary to correct both the low-frequency and high-frequency response of this type of amplifier if it is to be used for a very wide frequency band. The extent to which this correction is required depends on the characteristics of available amplifier tubes.

14.5. Vacuum Tubes Suitable for the Video Amplifier. It is evident from Eqs. 14.5 and 14.6 that the most important considerations in the choice of suitable tubes for the video amplifier are the transconductance and the capacity of the tube. An approximate criterion on which to base judgment of the merit of a tube for this purpose is the ratio of transconductance to capacity. However, this criterion must be used with caution because of the fact that the wiring capacity is added to the tube capacity, which decreases the dependence of gain on tube capacity alone.

Table 14.1 lists a few of the available types of tubes in order to estimate their performance in a resistance-coupled amplifier. Values are given for grid-to-cathode, plate-to-ground, and grid-to-plate capacities, and also for transconductance and plate resistance.

TABLE 14.1

Tube Type	Capacity, micro-microfarads			Trans-conductance, micromhos	Plate Resistance, ohms
	Grid-Ground	Plate-Ground	Grid-Plate		
6C5	4.0	13	2.0	2000	10,000
6C6	5.0	6.5	—	1200	1,000,000
954	3.0	3.0	—	1400	1,500,000
955	1.0	0.6	1.4	2000	12,500
1851, 1852	11.0	5.0	—	9000	750,000
1853	8.0	5.0	—	5000	700,000

The first tube listed is a triode. The total input capacity is, therefore, the grid-to-ground capacity plus the grid-to-plate capacity times the voltage gain of the tube. If the gain of this stage is large the input capacity is high owing to the reflected capacity. This excludes in general

the use of triodes in video amplifiers. An exception to this statement occurs in the first stage of a video amplifier, where, from the standpoint of signal-to-noise ratio, it may be desirable to use a triode. This will be discussed in greater detail in a following section.

The second tube, a pentode, is much more satisfactory. The total shunt capacity due to the tube electrodes is 11.5 micro-microfarads. This leads to a figure of merit of 1200/11.5, or, approximately, 100. The next tube, an acorn pentode, makes a very excellent video amplifier tube. Its figure of merit is about 230. There follows an acorn triode, the RCA 955. Here again the high reflected capacity limits its applicability in this type of amplifier. However, there are a number of positions in which it can play an important role because of its very low static capacity. The RCA 1851, 1852, and 1853 are specially developed television tubes. They are designed to meet the requirements of wide band amplification. A description of their characteristics is to be found in Chapter 17. The figures of merit of these tubes, as determined from the data given above, are:

RCA 1851, RCA 1852	550
RCA 1853	350

The calculated gain of a single stage will serve to indicate the performance of various members of this group of tubes. For this purpose Eq. 14.6 is used and a frequency f assumed such that f/f_0 is small compared to 1. The gain under these conditions will be merely Rg_m, where

$$R = \frac{1}{2\pi C_t f_0}, \qquad \text{so that} \qquad G = \frac{g_m}{2\pi f_0 C_t}.$$

Furthermore, a wiring capacity of 5 micro-microfarads has been added to the tube capacity to obtain the total capacity C_t. The results are tabulated in the following table, assuming f_0 to be 4.5 megacycles.

TABLE 14.2

Type	C_t	R	Gain
6C6	16.5 $\mu\mu$f	2100	2.5
954	11.0	3200	3.5
1852	25.0	1400	12.6

The superiority of the special television tube is very apparent, and

this tube, or its equivalent, is rapidly replacing all other types for the low level stages of video amplifiers.

14.6. High-Frequency Correction. The gain per stage of a simple resistance-coupled amplifier is given by Eq. 14.6. This equation shows that the gain has dropped to 70 per cent of its intermediate frequency value by the time the applied frequency has reached the characteristic frequency f_0. An amplifier with this characteristic would be serviceable only over a frequency range well below f_0. In order to accommodate the wide band required by the video signal, f_0 must therefore be very high. However, f_0 can be increased only by decreasing the coupling resistor R, since

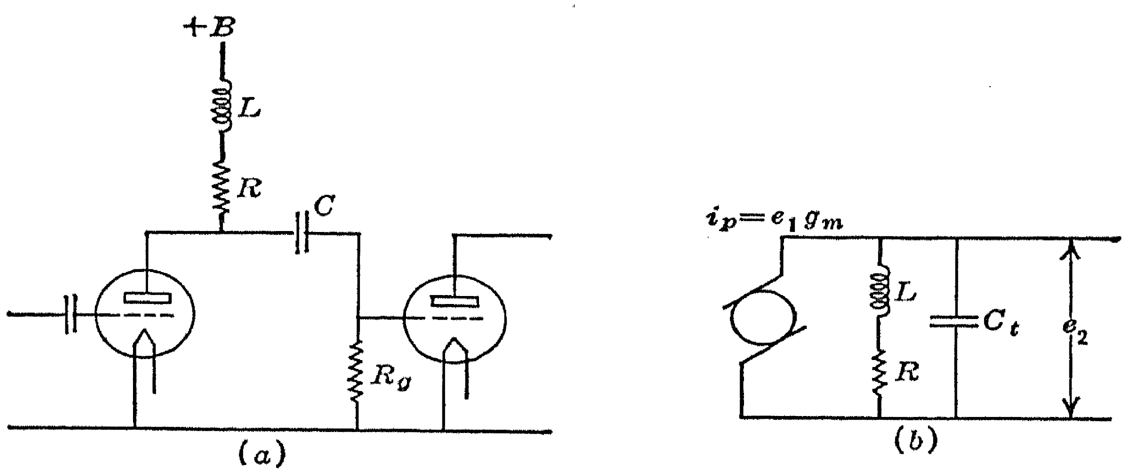


FIG. 14.9.—Response Correction by the Addition of an Inductance to a Resistance-Coupled Stage.

the shunt capacity is fixed. This, of course, decreases the intermediate frequency gain. It is obviously desirable to compensate the stage in order to improve its response at high frequencies.

The inclusion of a small inductance in series with the load resistance R is one of the simplest methods of obtaining partial correction of the response. This circuit, together with its high-frequency equivalent, is given in Figs. 14.9*a* and 14.9*b*. The voltage amplification for this stage is:

$$G = \frac{e_2}{e_1} = g_m|Z|, \tag{14.10}$$

where Z is the impedance of the coupling net at the frequency f. By ordinary circuit calculation methods this impedance is found to be:

$$Z = \frac{\dfrac{R}{2\pi f C_t} + j\left(\dfrac{L}{C_t} - 4\pi^2 f^2 L^2 - R^2\right)}{2\pi f C_t\left[R^2 + \left(2\pi f L - \dfrac{1}{2\pi f C_t}\right)^2\right]}. \tag{14.10a}$$

The absolute value of the gain is, therefore:

$$G = g_m \frac{\sqrt{\dfrac{R^2}{4\pi^2 f^2 C_t^2} + \left(\dfrac{L}{C_t} - 4\pi^2 f^2 L^2 - R^2\right)^2}}{2\pi f C_t\left[R^2 + \left(2\pi f L - \dfrac{1}{2\pi f C_t}\right)^2\right]}, \tag{14.11}$$

while the phase angle is given by:

$$\phi = \tan^{-1}\frac{\dfrac{L}{C_t} - 4\pi^2 f^2 L^2 - R^2}{R/2\pi f C_t}. \tag{14.11a}$$

At intermediate frequencies such that $1/2\pi f C_t \gg R$, the gain becomes:

$$G_i = g_m R.$$

As expressed in Eq. 14.11, the effect of the various circuit constants on the gain is not immediately apparent. A simplification can be effected by the substitution:

$$f_0 = \frac{1}{2\pi C_t R},$$

$$K = \frac{L}{C_t R^2}.$$

Rewriting Eqs. 14.11 and 14.11*a* in terms of f_0, K, and G_i, the gain as a function of frequency becomes

$$\frac{G}{G_i} = \sqrt{\frac{K^2\left(\dfrac{f}{f_0}\right)^2 + 1}{K^2\left(\dfrac{f}{f_0}\right)^4 - (2K - 1)\left(\dfrac{f}{f_0}\right)^2 + 1}}, \tag{14.12}$$

and the phase angle is:

$$\phi = \tan^{-1}\frac{f}{f_0}\left[(K - 1) - K^2\frac{f^2}{f_0^2}\right]. \tag{14.12a}$$

The value for K giving a gain ratio of unity at $f = f_0$ can be seen from Eq. 14.12 to be $K = \frac{1}{2}$. This is obtained by using the following circuit constants as derived from the definitions of f_0 and K:

$$R = \frac{1}{2\pi f_0 C_t},$$

$$L = \frac{1}{8\pi^2 f_0^2 C_t}.$$

The response under these conditions is not flat but rises to a value somewhat greater than unity at frequencies slightly below f_0 and then falls rather rapidly. The time delay is not constant, which is a more serious fault than the slight rise in the response.

More favorable response characteristics can be obtained with other circuit constants. The most nearly flat response is to be had if $K = 0.41$; $K = 0.32$ gives the most nearly constant time delay. In practice, a compromise value of K is generally used. Fig. 14.10 gives a family of response curves corresponding to various K values.

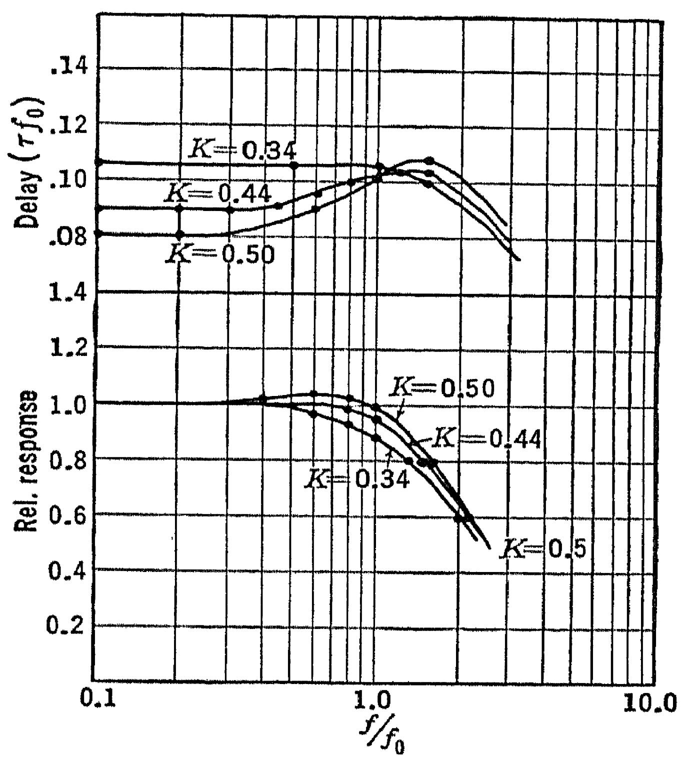


Fig. 14.10.—Families of Response Curves for Coupling Circuit Shown in Fig. 14.9.

14.7. The General Coupling Network. The resistance-coupled amplifier stage with a correcting element added to it is a special case of the general two-terminal coupling network. Although to date only a comparatively few members of this class of networks have been thoroughly examined, nevertheless valuable results have already been obtained and interesting generalizations have been reached.

An amplifier stage with a generalized two-terminal circuit network is shown in Figs. 14.11*a* and 14.11*b*. One limitation imposed on this network is that the tube capacities, C_t, must appear across the input of the circuit as shown in Fig. 14.11*c*. H. A. Wheeler* has shown, without rigorous proof, that, if Z is the most general two-terminal filter possible, the response giving a flat frequency characteristic over a bandwidth from a

* See H. A. Wheeler, reference 11.

very low frequency to the characteristic frequency f_0 is limited to $G = (g_m/\pi f_0 C_t)$. Actually this response cannot be fully realized in practice.

The response of the general circuit (Fig. 14.11a) is, of course:

$$G = g_m Z(f),$$

and its constancy is determined by the constancy of the impedance $Z(f)$. The general conditions of constancy for a given network can be easily

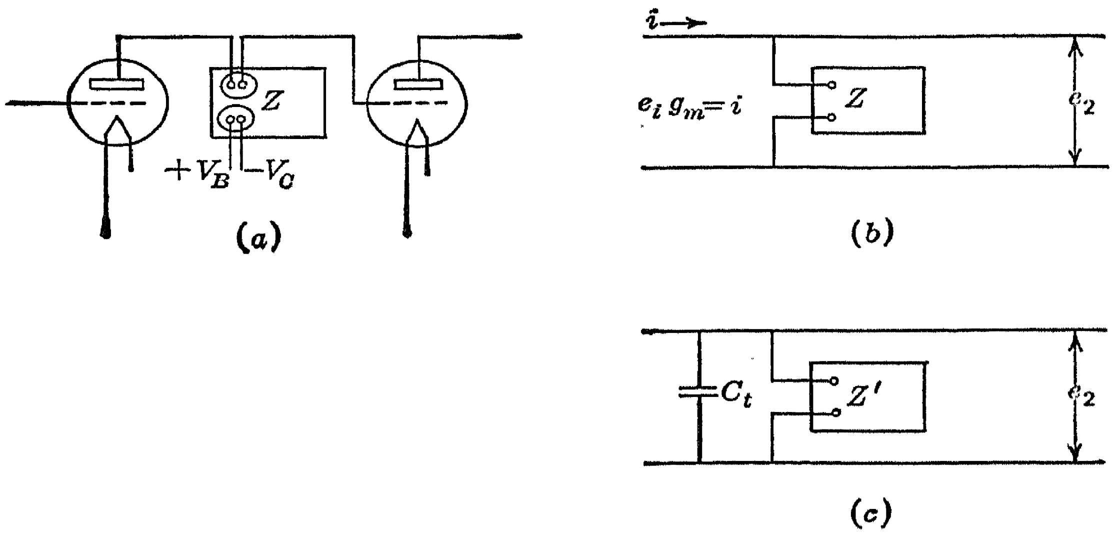


Fig. 14.11.—Two-Terminal Network as Amplifier Coupling.

determined.* The square of the impedance is expressed as a ratio of power series of the frequency, as shown:

$$|Z(f)|^2 = |Z(0)|^2 \frac{1 + a_1 f^2 + a_2 f^4 + \cdots}{1 + b_1 f^2 + b_2 f^4 + b_3 f^6 \cdots}. \qquad (14.13)$$

If now the coefficients of terms of like powers in the numerator and denominator are equal, the response is flat. Similarly, the derivative of the phase response can be expressed as a ratio of two polynomials, and a correspondence of coefficients of like terms is the criterion for a constant time delay.†

Referring again to the circuit discussed in section 14.6, it will be seen that Eq. 14.12 has been reduced to the type of fraction indicated by Eq. 14.13. The fact that the denominator contains $(f/f_0)^4$, whereas the numerator does not, shows that this type of coupling net cannot be made truly flat. However, for the range of frequencies over which $(f/f_0)^4$ may

* G. V. Braude showed that, if the impedance is developed as a Taylor's series and the coefficients are equated to zero, the response is flat. Tellegen and Verbeek originated the simpler method described, and showed it to be equivalent to Braude's method. *J. Tech. Physics, U.S.S.R.*, Vol. 4, Nos. 9–10.

† See E. W. Herold, reference 10.

be neglected in comparison to $(f/f_0)^2$, the amplifier response will be constant if

$$K^2 = 1 - 2K$$

or

$$K = 0.414.$$

If this value of K is introduced into Eq. 14.12, the amplitude response at $(f/f_0) = 1$ is found to be 92 per cent.

The phase angle for this circuit was given by Eq. 14.12a, which is:

$$\phi = -\tan^{-1}\left\{\left(\frac{f}{f_0}\right)\left[1 - K + K^2\left(\frac{f}{f_0}\right)^2\right]\right\}.$$

Differentiating Eq. 14.12a and equating coefficients, the condition for constant delay is fulfilled when

$$K^3 + 3K - 1 = 0,$$
$$K \cong 0.32.$$

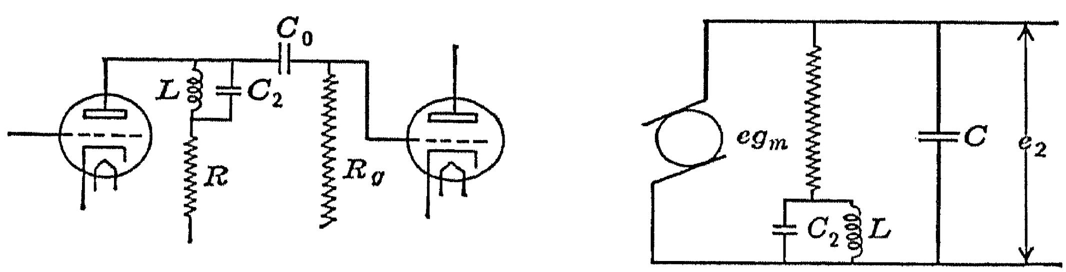


FIG. 14.12.—Coupling Network Including Inductance and Capacity for Response Correction.

It has been estimated that, if the amplitude response at the maximum frequency in the video signal does not fall below 90 per cent of the intermediate frequency gain, it may be considered satisfactory. If the value $K = 0.37$ is used, the amplitude response at $(f/f_0) = 1$ will be about 90 per cent of that at intermediate frequencies and the departure from constancy of time delay over this region will be 0.003 $(1/f_0)$. This is to be compared with a drop of the amplitude response to 90 per cent at $(f/f_0) = 0.48$ accompanied by 0.005 $(1/f_0)$ delay error for the circuit without the inductance (Fig. 14.4).

A condenser C_2 placed across the inductance, as shown in Fig. 14.12, increases the frequency over which the gain does not drop below 90 per cent. By the methods described above, it can be shown* that, for $K = 0.38$ and $C_2 = 0.30\ C$, the range of frequency is 1.25 f_0 and the time delay variation $0.008/f_0$.

* See E. W. Herold, reference 10.

A four-terminal network, as shown in Fig. 14.13, may also be used as interstage coupling. An advantage is gained in this case in that the tube capacities are not lumped but are at the two ends of the net. Wheeler has shown by very general methods that a maximum amplitude response of

$$G = \frac{g_m}{\pi f_0 \sqrt{C_p C_g}} \tag{14.14}$$

is obtainable with a flat characteristic when the impedances at the two ends of the net are properly matched to the respective capacities.

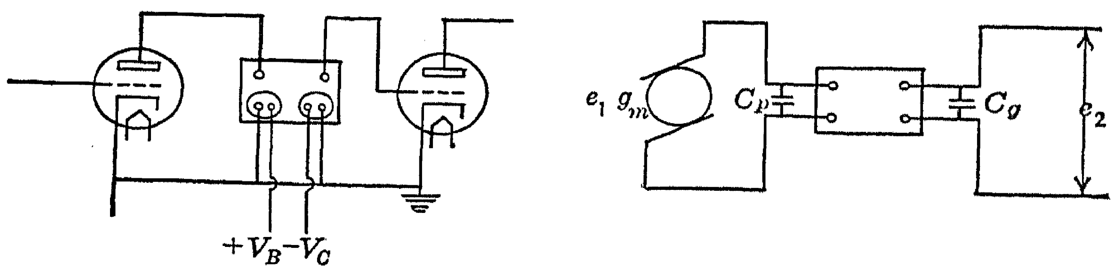


FIG. 14.13.—Four-Terminal Network as Amplifier Coupling.

A π network with an added inductance* in series with the load resistor, as shown in Fig. 14.14, represents a typical four-terminal coupling. The circuit constants giving flat response are $L_1 = 0.12\, CR^2$, $L_2 = 0.52\, CR^2$, $C_p = 0.34\, C$, and $C_g = 0.66\, C$. This circuit gives a 90 per cent response out to $(f/f_0) = 1.8$ and a delay error of $0.009\ (1/f_0)$.

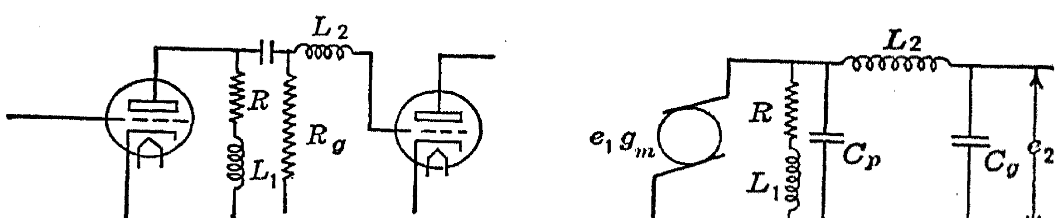


FIG. 14.14.—Simple Embodiment of the Four-Terminal Network as a Coupling Means.

The problem of interstage coupling from the standpoint of two-terminal and four-terminal filter theory has been investigated by Wheeler.† Considered from this point of view, the two-terminal coupling becomes a low-pass filter whose input impedance determines its effect on the response. A filter of constant-k sections, concluded with an m-derived half-section and giving an image impedance approximately matching the terminal resistor R (which in this case is the external plate resistor), is

* See G. V. Braude reference (p. 408); and E. W. Herold, reference 10.
† See H. A. Wheeler, reference 11.

shown in Fig. 14.15. Such a filter, if made up of a large number of sections, has an impedance given by

$$|Z| = \frac{R}{\sqrt{1 - (1 - n^2)\left(\frac{f}{f_0}\right)^2}}, \tag{14.15}$$

and a phase angle of

$$\phi = \tan^{-1}\frac{n\left(\frac{f}{f_0}\right)}{\sqrt{1 - \left(\frac{f}{f_0}\right)^2}}, \tag{14.15a}$$

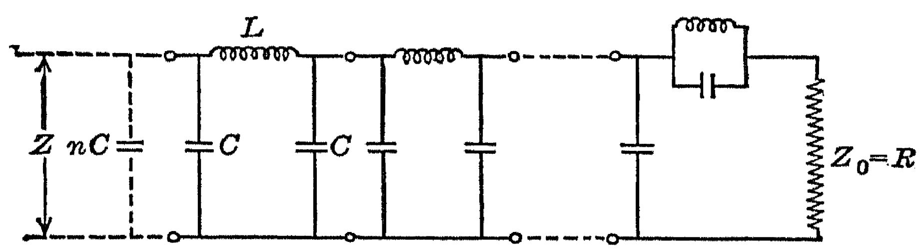


Fig. 14.15.—Constant-k Network.

where f_0 is $n/2\pi RC$. The total tube capacity in this case is $(1 + n)C = C_t$. The impedance for various values of n is shown in Fig. 14.16 together with the phase angle. The amplitude response becomes perfectly flat when n is equal to unity, and under these conditions the gain is:

$$G = g_m R = \frac{2g_m}{2\pi f_0 C_t}.$$

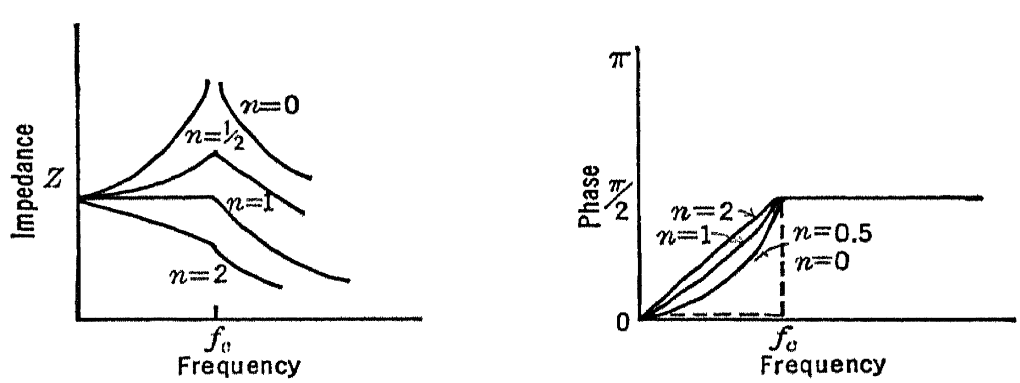


Fig. 14.16.—Characteristics of Network Shown in Fig. 14.15.

Under conditions of maximum constant gain, the time delay is not constant. In order to maintain an approximately uniform delay, n must equal 2, and, as is apparent from the form of the expression for Z, the response is not constant under these conditions.

It has also been shown that the three two-terminal coupling circuits which have been described above are elementary forms of this type of filter. With optimum circuit constants the circuit shown in Fig. 14.4 has an approximately uniform gain equal to 20 per cent of that of the ideal two-terminal network. The circuit of Fig. 14.9 is a single constant-k half-section and gives 55 per cent of the ideal response. Fig. 14.12 is an m-derived section and is capable of yielding 70 per cent. It is interesting to note that a circuit which includes a constant-k half-section and an m-derived half-section will give a gain which is 95 per cent of the theoretical maximum.

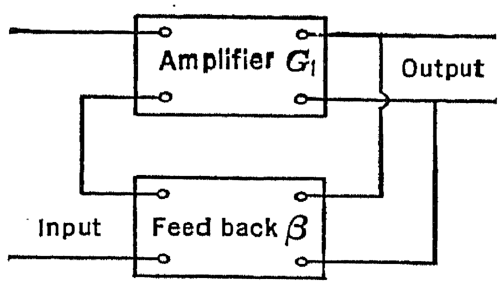


FIG. 14.17.—Block Diagram of Feedback Amplifier.

Although high theoretical efficiencies should be obtainable with these more complex circuits, and are to some extent realized in practice, difficulties such as those due to distributed capacity in coils and unwanted resistance losses are encountered in their physical embodiment.

Negative feedback may be used to improve amplifier response. This type of feedback consists of returning the output voltage to the grid through a suitable network, in such a way that it opposes the input voltage. Fig. 14.17 illustrates such a circuit schematically. If it is assumed that the gain of the amplifier is G_1 and the attenuation of the return network is β (either G_1 or β, or both, may be complex), the following relation exists between output voltage e_2 and input voltage e_1:

$$e_2 = G_1(e_1 - \beta e_2).$$

The net gain of the amplifier with feedback is, therefore:

$$\frac{e_2}{e_1} = G = \frac{G_1}{1 + G_1\beta}. \tag{14.16}$$

If, now, the voltage feedback is large compared with the input voltage, or, in other words, if the term $G_1\beta$ is large compared with 1, the net gain becomes:

$$G \cong \frac{1}{\beta}.$$

Thus the effective gain becomes independent of the properties of the amplifier and is determined solely by the characteristics of the return network. For example, if an amplifier has a gain of 2000 at intermediate

frequencies, and 1000 at the maximum frequency, and the loop has an attenuation of 0.01 independent of frequency, the net gain would be:

$$G = \frac{2000}{1 + 20} = 95 \text{ at intermediate frequencies,}$$

and

$$G = \frac{1000}{1 + 10} = 91 \text{ at maximum frequency.}$$

The effective gain is therefore nearly constant in spite of the 2-to-1 change in amplifier response.

Actually, because of the difficulty in stabilizing such an amplifier over the wide frequency band required for the video signal, it is not used in practice except in its simplest form. A frequently encountered simple feedback circuit is shown in Fig. 14.18. The feedback here cannot be made sufficient to render the net gain independent of the amplifier characteristics; consequently the return network in the cathode circuits is not made flat but is given properties compensating for the errors in the amplifier. The return net shown consists of a resistor shunted with a small condenser, making the negative feedback at high frequencies less than at low, compensating for the falling gain of the amplifier.

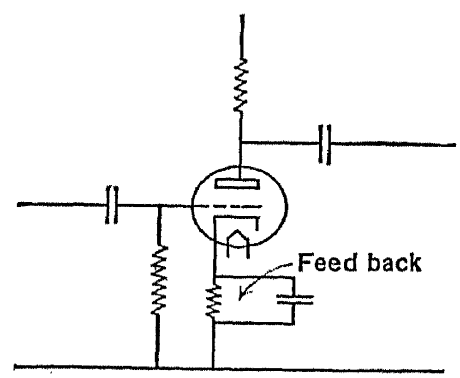


Fig. 14.18.—Simple Feedback Circuit for Correcting Frequency Response.

14.8. Low-Frequency Response. The television amplifier must not only respond to high frequencies, but must also be capable of amplifying signals down to below frame frequency of the picture being transmitted. The frequency range should theoretically extend to zero frequency if pictures of moving objects are to be perfectly transmitted. However, if an amplifier is good to a little below repetition frequency and very low frequencies corresponding to the average light in the picture are reinserted by methods which will be described later, the error introduced is negligible.

According to Eqs. 14.8 and 14.9 the expressions* for the response of a resistance-coupled stage at low frequencies are:

* These apply equally to the circuits discussed in sections 14.6 and 14.7 with the exception of those employing negative feedback.

$$G = g_m R_g R \frac{1}{\sqrt{R_g^2 + X_C^2}} = \frac{R g_m}{\sqrt{1 + \left(\frac{X_C}{R_g}\right)^2}},$$

$$\phi = \tan^{-1}\frac{X_C}{R_g},$$

$$X_C = \frac{1}{2\pi f C}, \qquad R \ll r_p \text{ and } R_g,$$

where R and R_g are the plate and grid resistors and C is the coupling capacity.

If the choice of C and R_g were not limited, the gain could be made constant down to any desired frequency. However, the choice of R_g is restricted by the grid characteristics of the tube in question, and in general cannot exceed about 500,000 ohms. The coupling condenser is limited by the fact that the physical size of a large capacitor introduces a large shunt capacity between grid and ground, and thus, as has been shown,

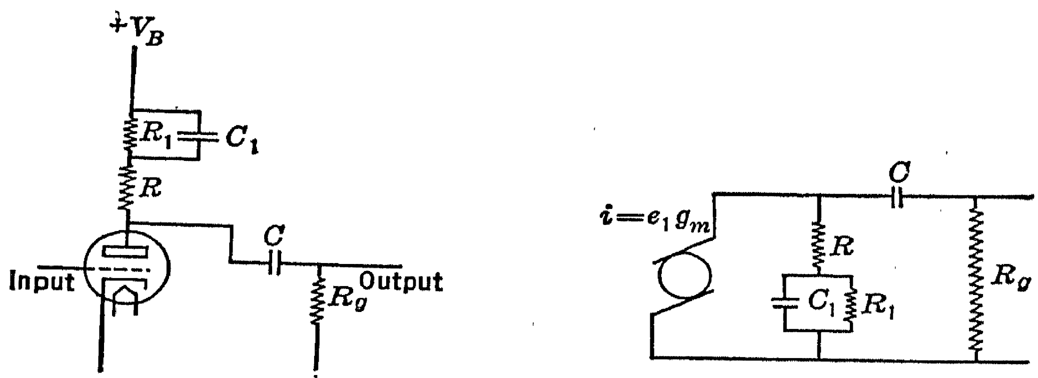


FIG. 14.19.—Circuit for Improving Low-Frequency Response.

has a very adverse effect upon the high-frequency response. A practical limit to the value of the coupling condenser may be taken as 0.1 microfarad.

Assuming a frame frequency of 30 cycles, the impedance of a 0.1-microfarad coupling condenser will be about 50,000 ohms. This means that the gain will be dropped ½ per cent of its value for intermediate frequencies. The phase delay, however, will be quite large, being about 10^{-3} second. In other words, a sine wave of this frequency impressed upon the grid will be delayed nearly 3 per cent of the vertical distance on the screen for each stage. This can, to some extent, be corrected by the filter shown in Fig. 14.19.

The correction factor to be applied to the gain as a result of the circuit illustrated is:

$$K = \frac{[(RR_1^2 + X_{c1}^2(R + R_1))^2 + (X_{c1}R_1^2)^2]^{\frac{1}{2}}}{RR_1^2 + X_{c1}^2 R}, \tag{14.17}$$

where $X_{c1} = (1/2\pi fC_1)$ and the term:

$$\phi' = \tan^{-1}\frac{X_{c1}R_1^2}{RR_1^2 + X_{c1}^2(R + R_1)} \tag{14.18}$$

is added to the phase angle.

As before, R is determined by the high-frequency response desired. This expedient does not yield complete correction, but, by properly choosing R_1 and C_1, the response at frame frequency can be made sufficiently good, both from the standpoint of phase and gain, for practical picture work.

Since this type of amplifier is made to respond to low frequencies, great care must be taken to decouple the common high-voltage supply, so as to avoid low-frequency relaxation oscillations or motor-boating. It is often found necessary to place a condenser of 1000 or more microfarads across the plate supply of such an amplifier.

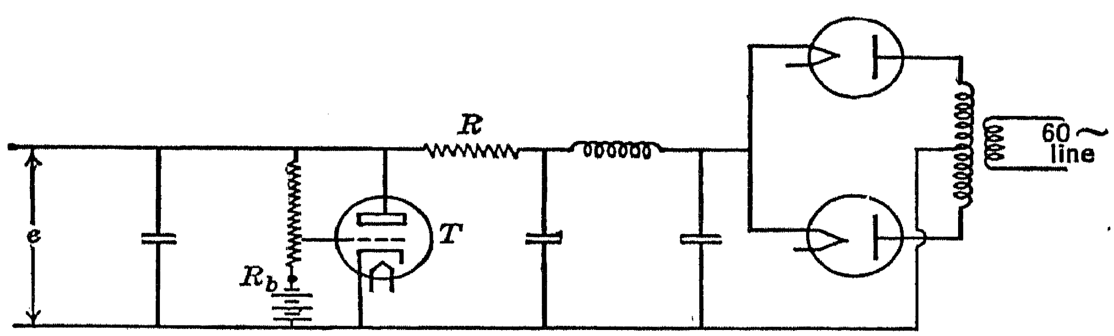


Fig. 14.20.—Regulated Voltage Supply.

Another method used, either alone or in conjunction with a large condenser, to lower the impedance of the voltage supply, is a vacuum-tube voltage regulator. One form of this regulator is shown in Fig. 14.20. The regulator tube T is arranged so that, as its plate current increases, the IR drop across the resistance R decreases the output voltage e. The grid potential with respect to the cathode is determined by the potential across R_b of the voltage divider. If now, for any reason, e tends to increase, the grid of tube T becomes more positive, causing the plate current to increase, which produces a drop along R opposing the increase of e. Greater sensitivity can be obtained if the single tube T is replaced by several stages of amplification. This type of constant-voltage plate supply is very useful for television work.

14.9. Overall Amplifier Response. The response of only a single stage of amplification has been considered so far. In practice, the complete amplifier consists of a great many cascaded stages, and the overall response is the important consideration. If the amplifier consists of identical stages and there is no accidental regeneration resulting from improper

shielding, the overall gain will be the gain of a single stage at each frequency raised to a power equal to the number of stages, while the time delay will be the product of the delay for each stage and the number of stages. The amplitude and delay curves* for 16, 32, and 64 stages of amplification coupled as shown in Fig. 14.9, with $K = \frac{1}{2}$ and $K = 0.44$, are given in Fig. 14.21.

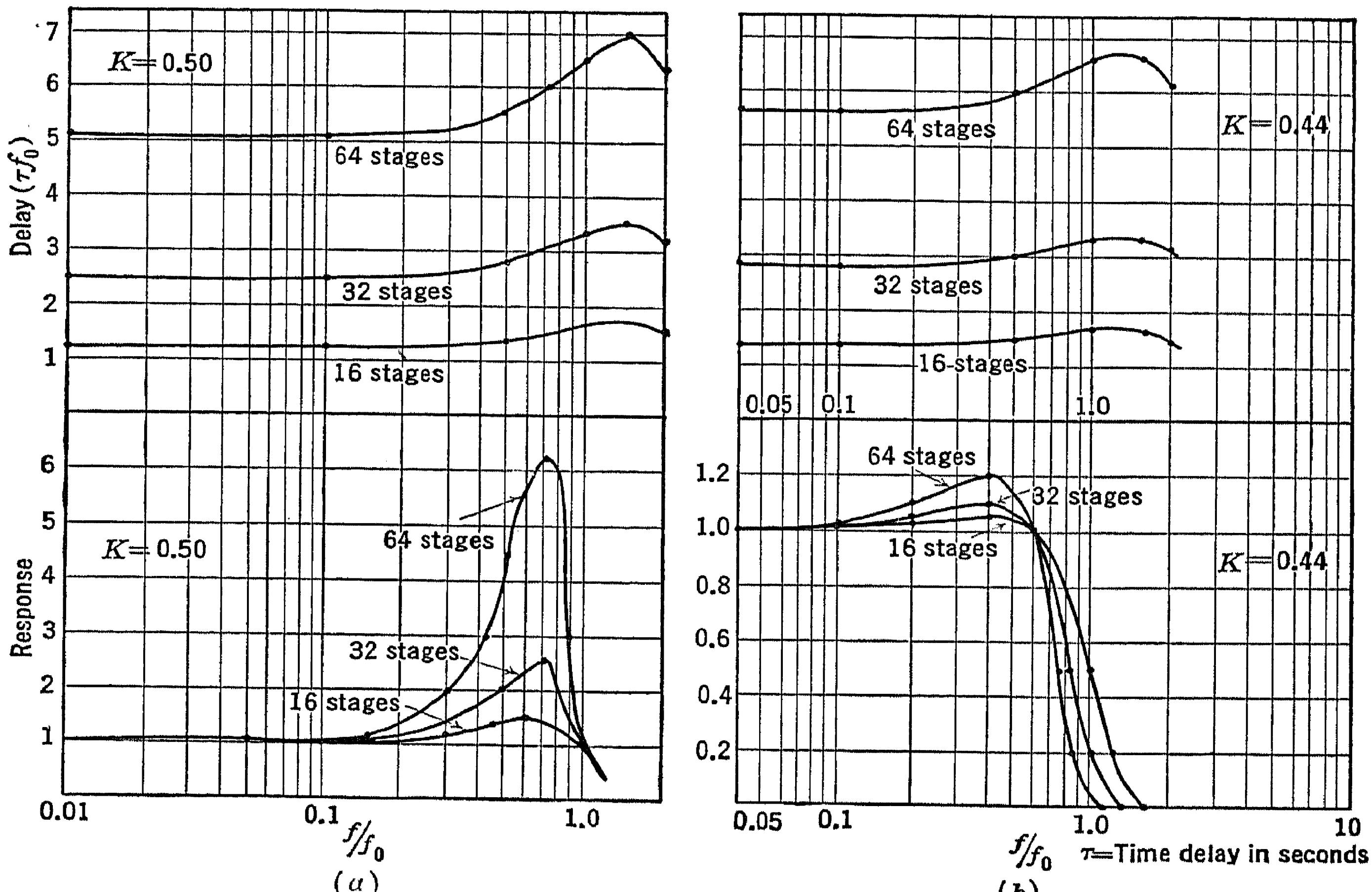


FIG. 14.21.—Response of Multi-Stage Amplifiers.

It is often advantageous to employ stages which are not identical but which differ in such a way that they tend to mutually compensate their individual defects. Thus, a filter-coupled stage which has a flat response but a positive time delay error over part of the useful frequency band can be used with another constant gain stage which has a negative delay error over the same portion of the band. If the characteristics of the individual stages differ, the overall gain is the product of the gains at each frequency, and the total delay error will be the algebraic sum of the individual errors.

In amplifiers having a great many stages, as at the pickup, monitoring, and modulating units of a transmitter, it is ordinarily not practical to make use of the complex coupling circuits required for high-gain stages

* See Bedford and Fredendall, reference 12.

throughout. This is so because of the extreme difficulty in maintaining constant gain and delay characteristics. It is general practice to use high-gain stages with response corrections as described for the first few stages —that is, until the signal level is well above the level of the noise generated by the amplifying tubes and coupling circuits. The remaining stages should be constructed to have a wide band at the expense of efficiency by using low-valued coupling resistors (increasing f_0), and require only a small amount of response compensation. The deviation from constancy of gain and time delay of the whole amplifier can thus be made to fall within acceptable limits.

14.10. Frequency Response Characteristics and Picture Quality. The analysis of the various amplifier circuits discussed in this chapter yields response characteristics in terms of amplitude and phase angle as func-

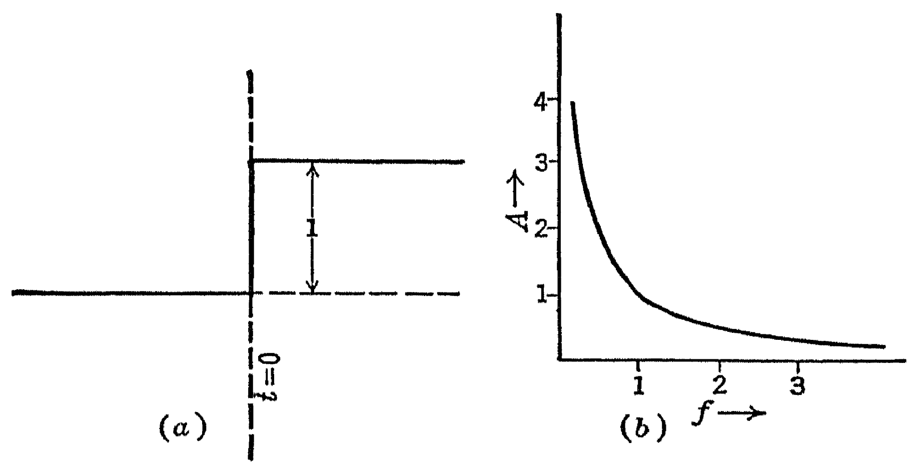


FIG. 14.22.—Frequency Distribution of Unit Voltage Step.

tions of frequency. It is difficult even after considerable experience to estimate directly from such response curves the quality of the picture that will be reproduced. Consequently, it is desirable to express the response characteristic in a form which can be more conveniently interpreted in terms of picture quality.

The response of an amplifier to a square voltage wave, or to a voltage step, is a very convenient indication of its behavior as a picture amplifier. The application of the gain and phase curves to determine the corresponding transient response involves the use of a Fourier integral. Furthermore, since graphical methods are the most convenient means of carrying out the transformation, it is expedient to separate the high- and low-frequency response.

The high-frequency transient response requires a knowledge of only the upper portion of the gain and phase curves, and assumes that from the intermediate frequencies down to zero frequency the gain and delay

are constant. On the other hand, the low-frequency analysis assumes that the upper end of the response spectrum is perfect.

Let it be assumed that the impulse shown in Fig. 14.22a is applied to the input of the amplifier. For convenience, the height of the voltage step is taken as unity. Therefore the input signal can be expressed as:

$$\begin{aligned} f_1(f) &= 0, \qquad -\infty<t<0. \\ f_1(f) &= 1, \qquad 0<t<\infty. \end{aligned}$$

The integral representation of this function is:

$$V = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin 2\pi ft}{f} df.$$

This can be interpreted to mean that the frequency spectrum of the voltage impulse applied to the input of the amplifier consists of continuous distribution whose amplitude is proportional to $1/f$. This distribution is shown in Fig. 14.22b. The fact that the amplitudes become infinite as f approaches zero causes no difficulty when considering a-c amplifiers. As far as the low-frequency portions of the response characteristics are concerned, the gain of the amplifier becomes zero before zero frequency. For the high-frequency analysis this portion of the spectrum is unimportant.

After passing through the amplifier the frequencies represented in the spectrum will be amplified by amounts which can be determined from the gain curves and will also be delayed by known phase angles. Consequently the output frequency spectrum will not be identical with the input spectrum. It will, however, represent some form of voltage impulse. The Fourier integral giving the form of this impulse is:

$$\begin{aligned} V = \frac{1}{2} &+ \frac{1}{\pi}\int_0^\infty \frac{A(f)\cos\phi(f)}{f}\sin 2\pi ftdf \\ &+ \frac{1}{\pi}\int_0^\infty \frac{A(f)\sin\phi(f)}{f}\cos 2\pi ftdf, \end{aligned} \tag{14.19}$$

where $A(f)$ is the gain of the amplifier as a function of frequency and $\phi(f)$ is the phase shift.

A graphical or numerical evaluation of this integral, though laborious, is not very difficult. $A(f)$, $\cos\phi(f)$, and $\sin\phi(f)$ can readily be found from the gain curves, while the functions $(\cos x/x)$ and $(\sin x/x)$ are available in standard tables.

To illustrate the information which can be obtained from a transient response analysis of an amplifier, Figs. 14.23a and b are given. Fig. 14.23a

shows the transient response curves for the three amplifiers having the frequency characteristics given in Fig. 14.21*a*, where $K = 0.5$; Fig. 14.23*b* shows similar curves for the response given in Fig. 14.21*b* for $K = 0.44$.* It is immediately apparent that, assuming the rest of the system to have an ideal response, the amplifier producing the curves in Fig. 14.23*a* cannot be used because of the extreme overshoot which would result in a dark line following every change from black to white in the picture, or a white line following a change in the opposite direction. The transient re-

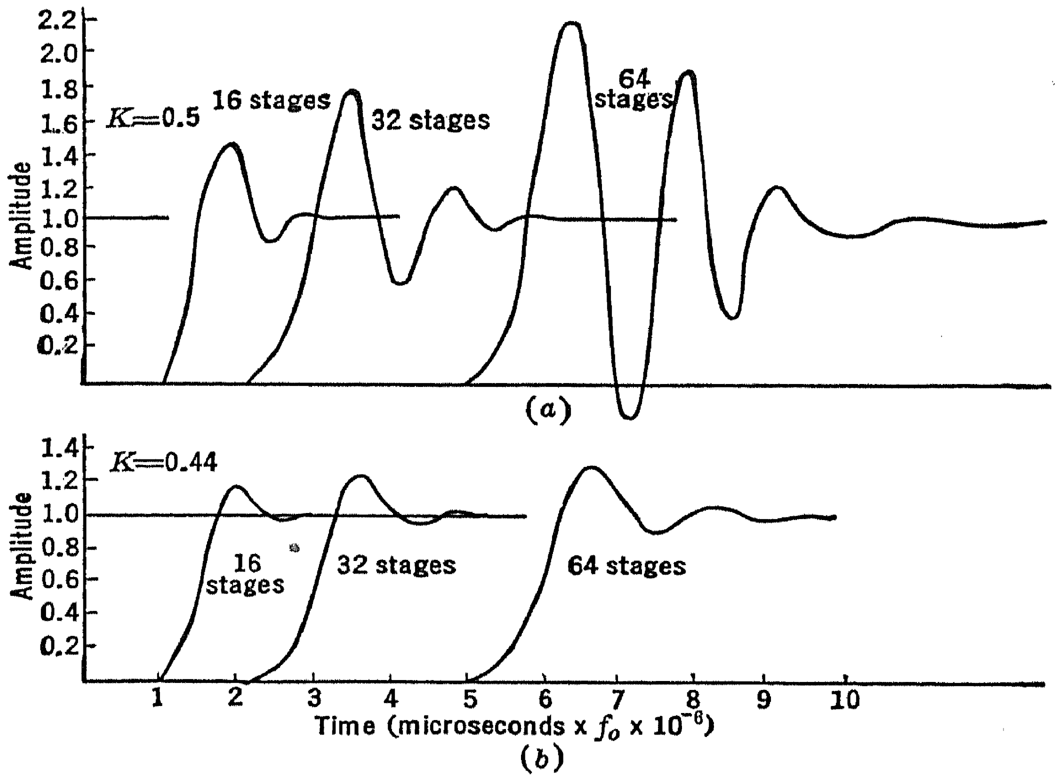


FIG. 14.23.—Transient Response to Unit Step of Amplifiers with Characteristics Shown in Fig. 14.21.

sponse shown in Fig. 14.23*b* will, however, be usable since the overshoot is only about 10 per cent, which is not objectionable. The "region of blurring" is greater for the second amplifier.

The analysis of the transient response to a unit voltage step may be extended to include the low-frequency portion of the amplifier spectrum. However, the calculations are laborious and the results not easily interpreted. The response of the amplifier to a square wave of frame frequency (representing a stationary horizontal bar) or of slightly less than frame frequency (a horizontal bar moving in the vertical scanning direction) yields more information. To illustrate the results of this type of analysis, Figs. 14.24*a*, *b*, and *c* show the characteristics of an amplifier which cuts off sharply at frame frequency, and its response to a square wave with

* See Bedford and Fredendall, reference 12.

the fundamental slightly less than frame frequency. The spurious waves which occur in Fig. 14.24c, representing the moving bar, are due entirely to the sharp cutoff. Any practical amplifier which has good response to frame frequency will have some response below frame frequency which will prevent these oscillations. If the phase delay departs seriously from linearity it will also result in waves or shading across the picture, a fact which should be evident without analysis.

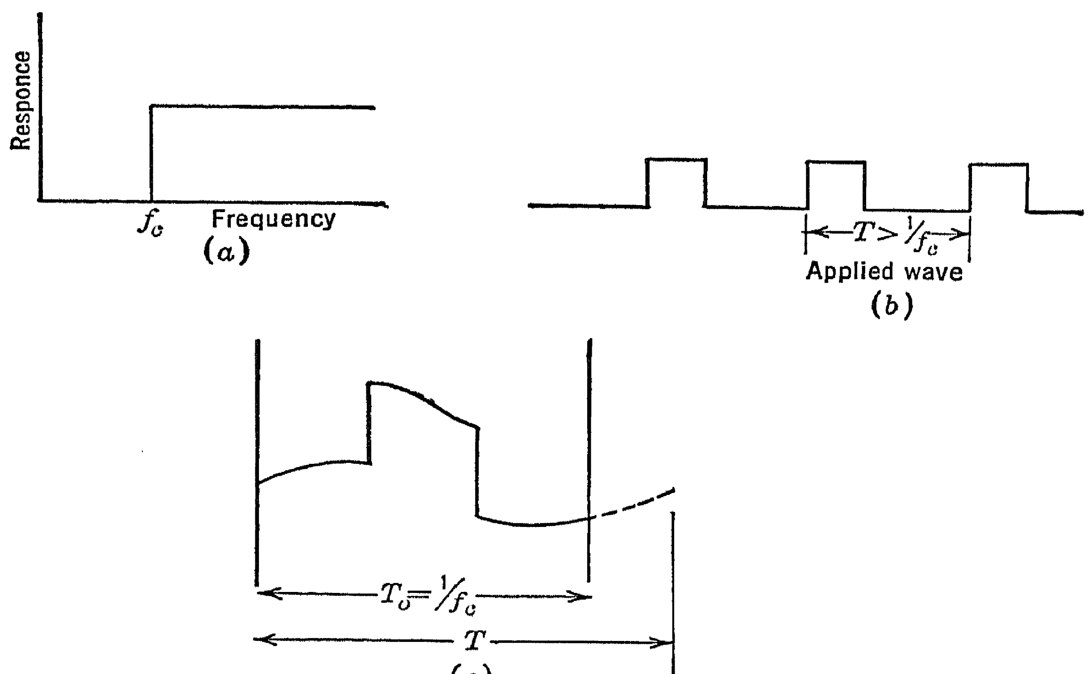


Fig. 14.24.—Effect of Low-Frequency Cutoff on Reproduced Signal.

14.11. Amplifier Response in Terms of Paired Echoes. The response of an amplifier of known amplitude and phase characteristics to a given input signal can be determined by evaluating the Fourier integrals given in the preceding section. This method is often very laborious, particularly if phase distortion is involved. H. A. Wheeler has proposed a very simple and ingenious procedure for estimating amplifier response which involves almost no calculation.* It can be used directly to determine the effect of small amounts of phase and amplitude distortion. By repeated application, or by a slight modification of the method, it may be used for any amount of phase or amplitude distortion. The method is based upon the interpretation of amplitude and phase distortion in the amplifier in terms of positive and negative pairs of echoes.

The principle involved in the method is apparent from the following consideration: Assume that a network has applied to its input the impulse shown in Fig. 14.25a. The network is of such a character that the main pulse is transmitted through without distortion, appearing at the output in time t_0, and that accompanying the main pulse are echoes

* See Wheeler, reference 13.

which precede or follow by a time t_e. These echoes are identical in shape with the main pulse but are attenuated by a factor a. The resultant pulse at the output is the sum of the main pulse and its echoes. Fig. 14.25b shows the main pulse at time t_0 and two echoes, one preceding, the other following, at times $t_0 - t_e$ and $t_0 + t_e$. The resultant pulse is the sum of

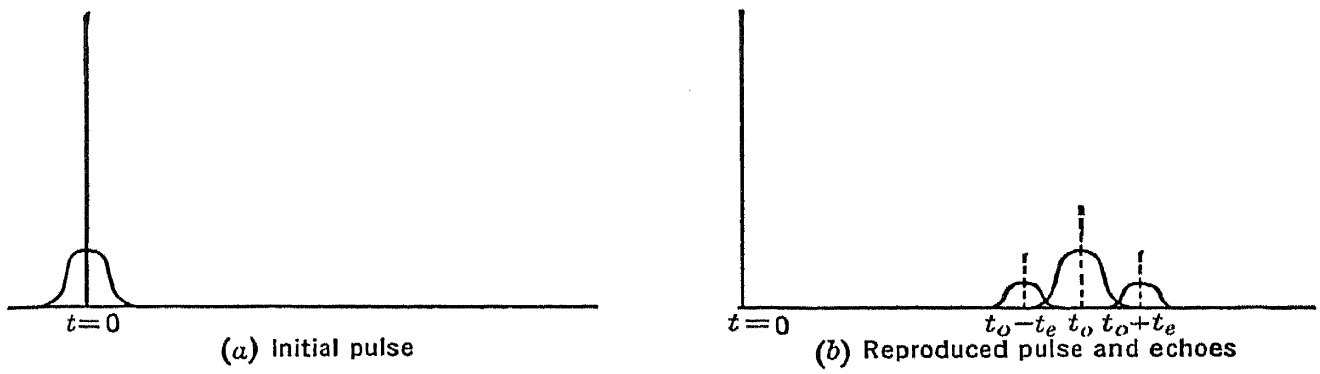


Fig. 14.25.—Echoes Accompanying Transmitted Pulse.

these curves. The pulse and echoes may be resolved into their frequency components, and the spectral distributions will be found to be identical except for the attenuation factor a reducing the amplitude of the echo frequencies. Since the echoes are displaced in time with respect to the main pulse by a constant amount t_e, there will be a phase difference β

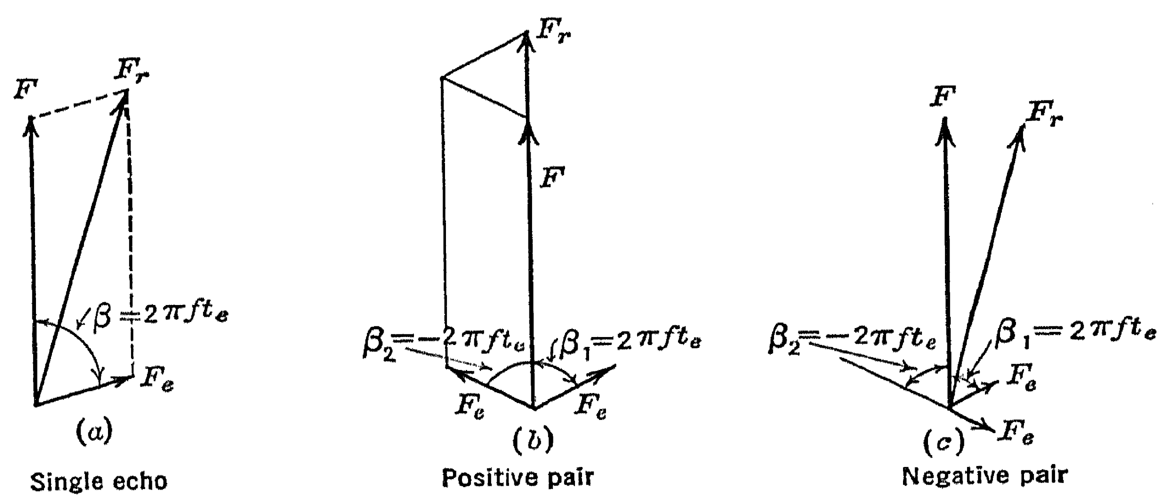


Fig. 14.26.—Vector Addition of Main and Echo Frequency Components.

between any frequency f of the main pulse and the corresponding frequency in the echo, which is given by

$$\beta = 2\pi f t_e.$$

The amplitude and phase of any component in the spectrum of the resultant pulse may be considered as the vector sum of the corresponding components in the main pulse and the echoes. This addition for a single echo is shown in Fig. 14.26a. It will be noticed that both amplitude and

phase of the resultant differ from those of the frequency component of the original pulse. Two special types of echo groups are of particular interest.

One of these groups is a pair of echoes of equal magnitude which lead and follow the main pulse by equal amounts of time, i.e., $+t_e$ and $-t_e$. This is shown in Fig. 14.25*b*. A group of this type has been termed a positive pair of echoes. Under these conditions, when the pulse and echo frequency components are added together there is no phase difference between those of the main pulse and the resultant. The only effect is a difference in amplitude as will be apparent from Fig. 14.26*b*. If the amplitude of the main pulse frequency is F and that of each echo aF the amplitude of the resultant is:

$$F_r = F(1 + 2a \cos 2\pi f t_e).$$

If, instead of a single positive pair of echoes, there are a number of pairs at times $\pm t_e, \pm 2t_e, \pm 3t_e \cdots \pm nt_e$ with size factors $a_1\, a_2\, a_3 \cdots a_n$ the resultant amplitude of any frequency f is:

$$F_r = F(1 + 2a_1 \cos 2\pi t_e f + 2a_2 \cos 4\pi t_e f + 2a_3 \cos 6\pi t_e f \cdots + 2a_n \cos 2n\pi t_e f). \qquad (14.20)$$

It is often convenient to relate the shortest echo time t_e to a nominal cutoff frequency f_c by the following relation:

$$t_e = \frac{1}{4f_c}.$$

Hence Eq. 14.20 may be written as:

$$F_r = F(1 + g(f)), \qquad (14.21)$$

$$g(f) = \sum 2a_k \cos \frac{k\pi f}{2f_c}. \qquad (14.21a)$$

Since Eq. 14.21*a* has the form of a typical Fourier series, the resultant spectrum F_r may have any shape between 0 and the cutoff frequency. Conversely, if the spectrum of F is known and that of the resultant pulse F_r, the coefficients a_k can be evaluated. Finally, if the ratio $(F_r/F) = G$ is formed, it will be seen to be the amplitude response of the network, and from it alone the coefficients of the series expressing $g(f)$ can be determined. Therefore, if the response of an amplifier to a given pulse is desired, it is only necessary to determine the coefficients a_k giving the magnitude of the echoes which result in an amplitude characteristic corresponding to that of the amplifier. The main pulse and echoes can then be drawn on a time scale and the resultant output pulse constructed graphically.

Usually the amplitude response characteristic departs very greatly from the ideal flat response, particularly at high frequencies. If the coefficients are determined from a comparison of the actual response with the ideal, a great many coefficients will be required and the echo pattern will be very complicated. To avoid this, the amplification process may be considered as occurring in two steps. For example, if the response to the rectangular impulse shown in Fig. 14.27*a* by an amplifier whose characteristic is shown by the curve G in Fig. 14.27*b* is to be determined, it may be considered as first passing through a network which has a transmission characteristic G_1 shown in Fig. 14.27*d*, then through one having the characteristic G/G_1 shown in Fig. 14.27*f*. The characteristic G_1 is so chosen that its response to the test pulse can easily be determined by ordinary methods. Fig. 14.27*c* shows the form of the pulse after having passed through the first network. The second part of the amplifier representation, which has the transmission characteristic G/G_1, can, in the example illustrated, be adequately expressed by retaining a single coefficient in the series $g(f)$. Its effect on the output pulse is to produce the single pair of echoes which have a polarity opposite from that of the main pulse. The resultant output from the overall amplifier when the rectangular impulse is applied to the input is the sum of the main pulse and the echoes, as shown in Fig. 14.27*e*. The amplitude characteristic G_1 is so chosen that it can be used for a number of amplifiers and the resulting output (*c*) determined once and for all.

The second type of echo group also consists of a pair of echoes which lead and follow the main pulse by equal time intervals. However, the two echoes are opposite in polarity. Such a group has been termed a negative pair of echoes. Considering a single frequency of the main pulse and the echoes, it will be seen from the vector diagram in Fig. 14.26*c* that to a first approximation their effect is to change the phase of the resultant signal without altering its amplitude. If

$$F \cos (2\pi ft + \phi_0)$$

is a frequency component of the main pulse, the resultant frequency component will have the following form:

$$F \cos (2\pi ft + \phi_0 + 2b \sin 2\pi ft_e),$$

where b, which is assumed to be small compared to unity, is the ratio of amplitude of the echo to the main pulse.

If there is a series of negative pairs of echoes at intervals t_e, $2t_e$, $3t_e$ $\cdots$ nt_e, the resultant components can be represented by:

$$F \cos (2\pi ft + \phi_0 + \sum 2b_k \sin 2\pi kft_e). \tag{14.22}$$

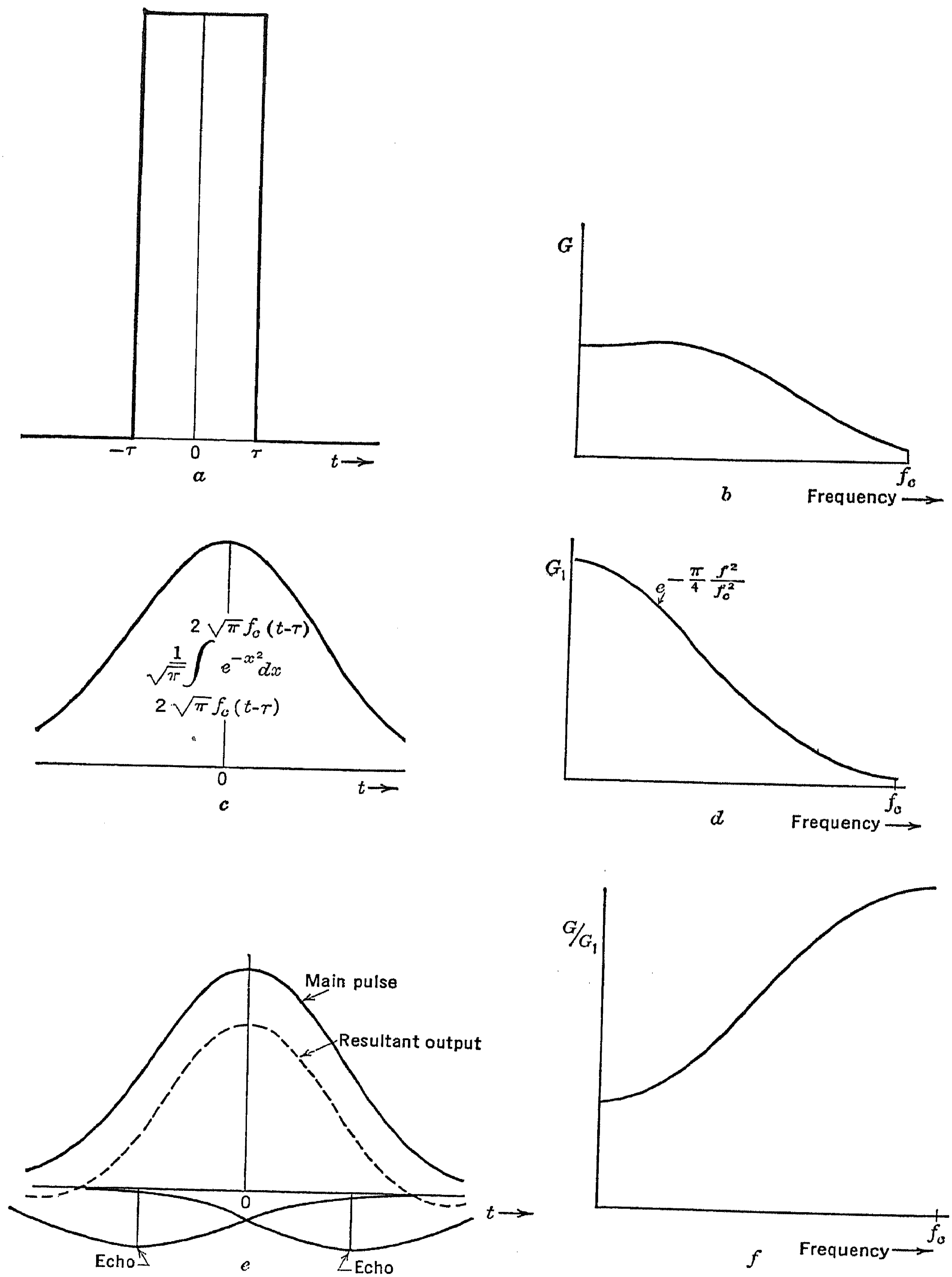


FIG. 14.27.—Determination of Resultant Pulse in the Presence of Amplitude Distortion with the Aid of Positive Paired Echoes.

As has already been pointed out, if the phase of the input frequency component is taken as a reference, a phase shift through the amplifier of

$$\phi_0 = b_0 f$$

produces no distortion, but results only in a time delay $b_0/2\pi$ of the signal. Therefore, of the total phase shift

$$b_0 f + \sum 2b_k \sin 2\pi k f t_e, \tag{14.23}$$

only the series represents a distortion. This distortion of the output can be regarded as the result of a group of negative echo pairs whose sizes are given by the coefficients b_k. The determination of the distortion resulting from a phase characteristic which is not linear with frequency is

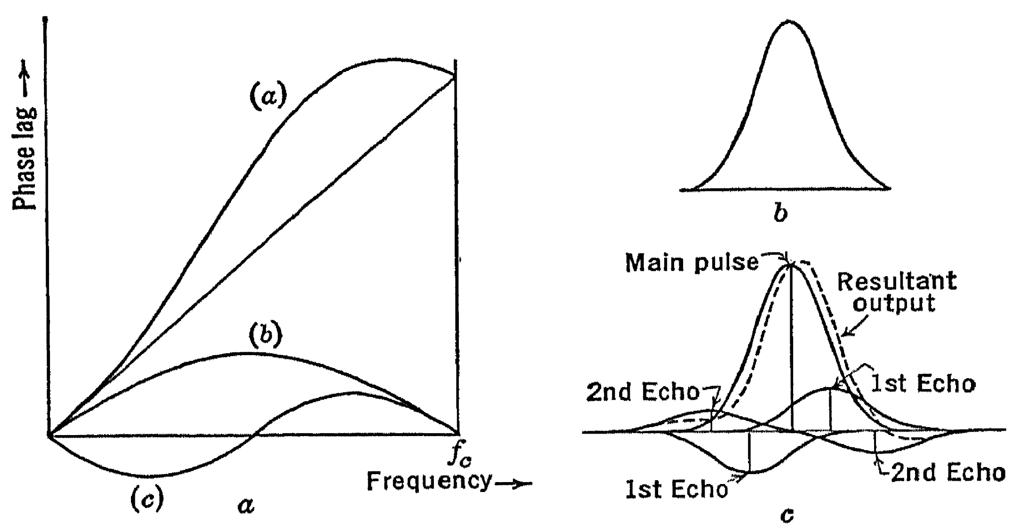


Fig. 14.28.—Phase Distortion in Terms of Negative Paired Echoes.

illustrated by the following example. The amplifier is divided into an element which has no phase distortion, and one which has no amplitude defect but a phase characteristic as represented by curve (a), Fig. 14.28a. The impulse from the first element is assumed to have been determined by the methods previously described. The departure from a straight-line characteristic can be resolved into two components, shown in curves (b) and (c) of Fig. 14.28a, corresponding to two terms, $2b_1 \sin (\pi f/2f_c)$ and $2b_2 \sin (2\pi f/2f_c)$ of Eq. 14.23. The phase distortion can be interpreted as two negative echo pairs as shown in Fig. 14.28c, giving the resultant pulse shown in the same figure.

This interpretation of amplitude and phase distortion in terms of echo pairs permits a quick estimate of the effect of a given video amplifier characteristic on the appearance of the reproduced picture.

14.12. Additional Frequency Correction. All pickup devices at present in use in practical television systems give a response which is more

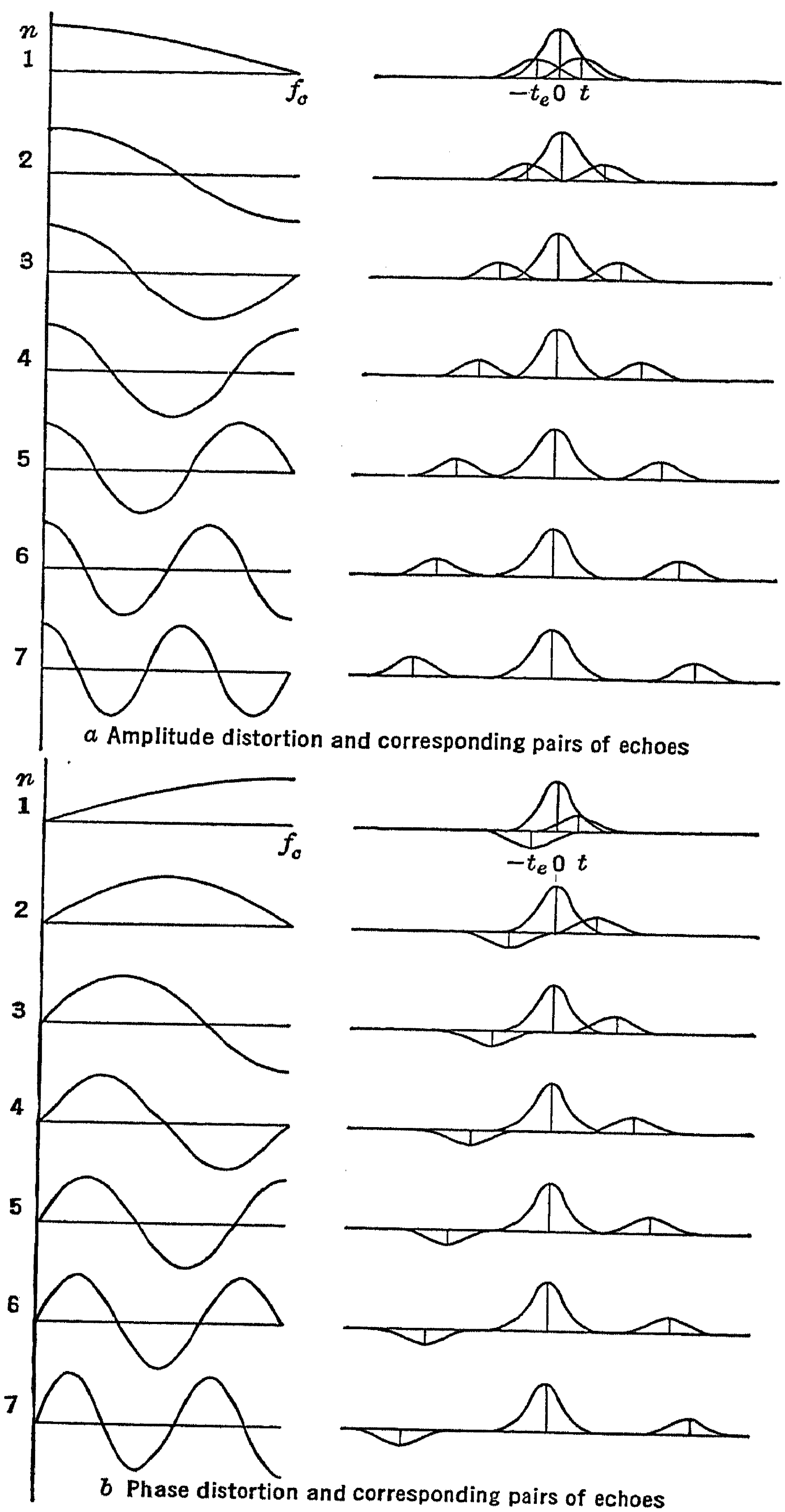


FIG. 14.29.—Echo Pairs for Various Distortion Characteristics.

or less dependent on frequency. This is due in part to the electrical characteristics of the device, and in part to aperture defect. In order to obtain undistorted image reproduction it is necessary to correct the amplifier circuits in such a way that the overall response is flat.

Such compensation is necessary with the Iconoscope. The electrical characteristics of this tube are those of a constant current generator shunted by a capacitance or, more exactly, a voltage source in series with a resistance of several megohms and a condenser of about 10 micromicrofarads capacity across the output. The capacity is, of course, in parallel with the resistance coupling the tube to the video amplifier. Since the coupling resistance must be high in order to maintain a high signal-to-noise ratio there will be an attenuation of the signal at high frequency.

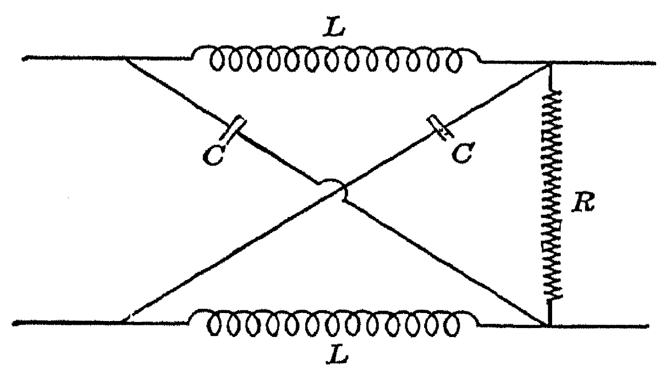


Fig. 14.30.—Phase Correcting Network.

Scanning with a finite aperture always results in a decrease in signal at the higher frequencies. If the scanning aperture is very small the frequency distortion produced may be negligible over the working range of the video amplifier. However, in most practical pickup devices, including the Iconoscope, aperture defect cannot be neglected. It has already been shown in Chapter 6 that the effect of the aperture on the signal is equivalent to that of a filter in series with an ideal signal generator. This equivalent filter has an attenuation which increases with frequency, but produces no change in phase delay.

To compensate for the impedance characteristics and aperture defect, a circuit network with a frequency attenuation characteristic corresponding to the inverse of that of the pickup device must be provided in the amplifier. This network is not necessarily separate from those used to compensate the frequency response of the amplifier itself, and, in fact, the correction is often made by over-compensating one or more of the amplifier stages with the circuits described in the preceding sections. Occasionally, however, it is advisable to use a separate filter for this correction, and in this case a high-pass filter (either dissipative or non-dissipative) can be designed on the basis of circuit theory to fit the attenuation requirements.

Since the attenuation due to aperture defect occurs without phase changes, and most filters used to compensate for this attenuation introduce a degree of non-uniformity in the delay characteristics, a phase correcting network may be necessary. In general, these nets are derived

from the basic lattice network shown in Fig. 14.30. If the constants of such a network are so chosen that:

$$R^2 = \frac{L}{C},$$

the attenuation will be constant while the phase which is given by

$$\theta = \tan^{-1} \frac{2R}{\left(2\pi f L - \frac{1}{2\pi f C}\right)} \tag{14.24}$$

is a function of frequency. The phase angle as a function of f/f_0, where $f_0 = 1/(2\pi\sqrt{LC})$, is given in Fig. 14.31. Other types of phase variation can

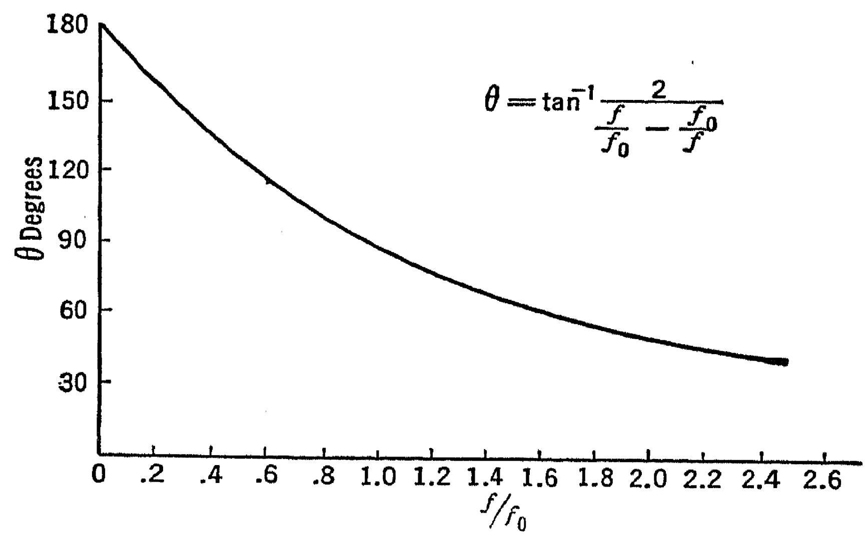


FIG. 14.31.—Phase Response of Network Shown in Fig. 14.30.

be obtained by substituting various combinations of inductance and capacitance in the bridge arms.

With the aid of either the amplifier compensating circuits or, where necessary, a high-pass filter, and a phase compensating net, it is possible to meet the requirements of overall uniformity in amplitude and delay as a function of frequency.

14.13. Non-Linear Amplification. All vacuum-tube amplifiers have a non-linear voltage response. The non-linearity can be made very small if the voltage swing on each stage is kept within the so-called linear portion of the tube characteristic, but it can never be eliminated. As has been seen from the foregoing chapters, both the Iconoscope and the Kinescope are likewise non-linear in their response. As a result the overall amplitude characteristic of the complete system, also, is invariably non-linear. The effect of this non-linearity is to distort the contrast values in the repro-

duced picture, as has already been pointed out in Chapter 6. The response of a typical Iconoscope-Kinescope system to a series of bright bands having equal increments of light intensity is shown in Fig. 14.32.

Amplitude distortion can be compensated to some extent by an added non-linear stage in the amplifier, but this expedient is rarely found necessary in practice. Fortunately, the eye is accustomed to seeing under such a wide range of contrast conditions that a contrast distortion many times that ordinarily found in the reproduced picture would be entirely unnoticeable. In this respect, the eye differs greatly from the ear, which is very sensitive to non-linear distortion. The amount of distortion of this type existing in even the best television system could not be tolerated in an audio system.

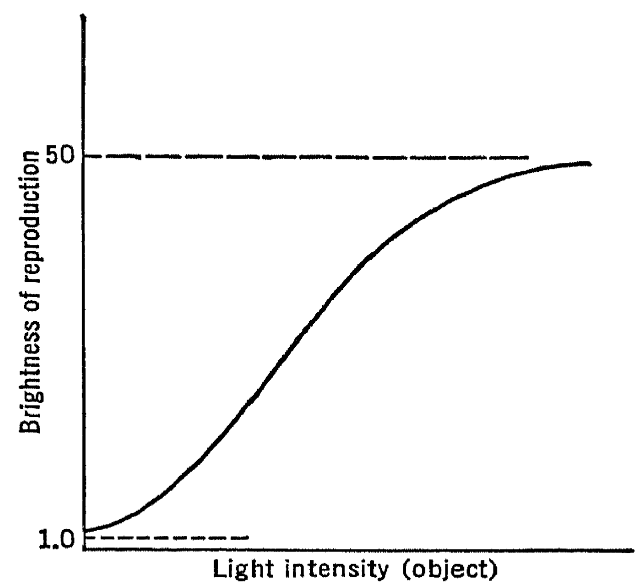


Fig. 14.32.—Non-Linearity of a Typical Iconoscope-Kinescope System.

14.14. Noise Considerations. The importance of noise in relation to the picture signal has already been mentioned in the discussion of the operation of the Iconoscope. It has been shown that the limit to the sensitivity of the pickup tube is determined by the noise generated either by the tube itself or by the amplifier. Observations were cited in Chapter 10 which indicated that for a completely satisfactory picture the ratio of the voltage generated by the highlights and the root mean square noise voltage must be at least 30; in other words, a signal-to-noise ratio of 30 or better is required. It was also pointed out that a picture having a signal-to-noise ratio of 10 could be considered usable.

The basic noise in a television system is generated by the pickup tube and the first few amplifier stages. Additional noise may be introduced at other points in the system as, for example, at the input of the radio receiver used to pick up the television signal, but this added noise is not fundamental and may be eliminated by improving the receiving antenna, increasing the transmitter power, etc.

Where an Iconoscope is used as the pickup device the noise generated by the tube itself is determined primarily by the current in the scanning beam. By using a low beam current the noise generated by the Iconoscope can be reduced to such a value that it can be neglected in comparison with that produced by the amplifier. With a normal Iconoscope there

is an optimum beam current for which the magnitude of the picture signal is a maximum. If the tube is operated below this optimum, signal output is reduced, but there is an even greater decrease in spurious signal, so that the overall picture is improved if the noise level of the amplifier permits operating in this range. The lower limit beyond which the beam current cannot be reduced when using a normal Iconoscope is determined by the current required to discharge the mosaic completely. If this minimum beam produces a noise which is much in excess of the amplifier noise (a condition never realized with the present thermionic amplifiers), a further decrease in beam current without loss in efficiency can be effected by reducing the capacity per unit area of the mosaic, as was explained in Chapter 11.

There are two fundamental sources of noise in a vacuum-tube amplifier after such avoidable sources as poor connections or insulation, bad batteries, microphonics, and external pickup have been eliminated. These are the noise generated by the thermal agitation in the resistors and the emission noises in the vacuum tubes.

14.15. Thermal Noise of Resistors. It has been shown experimentally by Johnson*—after whom the effect has been named—and theoretically by Nyquist* that a resistor, regardless of the material of which it is made and independent of the current flowing through it, generates an emf across its terminals owing to the fact that its conduction electrons are in thermal equilibrium with the atoms. It can be shown that this noise is made up of purely random fluctuations, and that, like shot noise discussed in Chapter 1, the noise power is distributed uniformly over the entire frequency band. Furthermore, the mean square voltage is proportional to the resistance of the resistor or, in the case of a complex impedance, to its resistive component. Finally, the power is found to be proportional to the absolute temperature of the resistive element.

The relation between the voltage generated and the above-mentioned factors can be formulated as follows:

$$\overline{E^2} = 4kRT(f_2 - f_1), \tag{14.25}$$

where k is Boltzmann's constant; T, the absolute temperature; $\overline{E^2}$, the mean square voltage; f_2 and f_1, the upper and lower limit, respectively, of the frequency band over which measurements are being made; and R the resistance.

At room temperature this expression becomes:

$$\overline{E^2} = 1.6 \times 10^{-20} R(f_2 - f_1) \text{ volts}^2. \tag{14.25a}$$

* See Johnson, reference 15; and Nyquist, reference 16.

14.16. Tube Noise. Three types of noise can occur in vacuum tubes. These are flicker effect, shot noise, and space-charge-limited noise. In addition, there is noise due to ions, secondary emission, and leakage, but in a well-designed tube used for television purposes these last are negligible.

Flicker effect is caused by changes in emission over small parts of the cathode, apparently due to chemical and crystal-structure changes of the emitter. The fluctuations resulting from this effect are important over a rather narrow frequency band at the lower end of the spectrum and become negligible above 1000 cycles. In general, it is not a serious obstacle in television work.

Shot noise has been discussed in detail in Chapter 1. In a temperature-limited vacuum tube this type of noise predominates over all other tube noises. It can be quantitatively expressed in terms of the emission current I by the following equation:

$$\overline{i^2} = 2e(f_2 - f_1)I. \tag{14.26}$$

When a tube is operated under space-charge-limited conditions, there is a reaction between the fluctuation in emitted electron current and the height of the potential barrier at the virtual cathode which reduces the effect of shot noise in the anode current. This does not mean, however, that the current output from the tube will be free from statistical fluctuations. Theoretical investigation* of this effect has led to a quantitative explanation under certain operating conditions. An approximate rule applicable to screen-grid tubes states that the mean square fluctuations in the plate current will be equal to the temperature-limited shot noise of a current equal to that reaching the screen.† In general, triodes exhibit a lower noise level than the equivalent multi-grid tubes.

Tubes are frequently rated for their noise characteristics in terms of an equivalent noise resistance, which is determined as follows. The mean square plate current is divided by the square of the transconductance. This yields the mean square voltage which, impressed on the grid, would produce the same noise output in an identical "noiseless tube." The equivalent noise resistance is the resistance whose thermal noise voltage at room temperature is equal to this voltage. This gives a rating which is independent of the frequency band involved and the gain of the tube. For example, if upon measuring the plate current fluctuations of a 57-type tube, whose transconductance is 1200 micromhos, over a 10-

* See Thompson and North, reference 19; and Percival and Horwood, reference 20.

† The rule can be applied only to tubes in which the screen-grid current is large; it is not applicable to such tubes as the RCA 6L6.

kilocycle band, the noise is found to be 3.4×10^{-18} ampere^2, the rating in terms of equivalent ohms across the grid will be:

$$\overline{V^2} = \frac{\overline{i^2}}{g_m^2} = 2.4 \times 10^{-12} \text{ volt}^2,$$

$$R_t = \frac{\overline{V^2}}{1.6 \times 10^{-20} \times F} = \frac{2.4 \times 10^{-12}}{1.6 \times 10^{-20} \times 10^4} = 15{,}000 \text{ ohms.}$$

For screen-grid tubes falling within the scope of the rule given above the equivalent noise resistance will be given by:

$$R_t = \frac{32 \times 10^{-20} i_s \Delta f}{1.6 \times 10^{-20} \Delta f g_m^2 10^{-12}} = \frac{2 \times 10^{13} i_s}{g_m^2}, \tag{14.27}$$

where i_s is the screen current and g_m is the transconductance in micromhos.

The noise ratings* of a few characteristic tubes are listed in Table 14.3.

TABLE 14.3

Tube Type	Noise Rating R_t	Mode of Operation
57	15,000 ohms	Pentode
6C6	6,000	Pentode
6C6	1,600	Triode
954	5,500	Pentode
955	1,500	Triode
6L6G	1,050	Tetrode
6L6G	600	Triode
1852	520	Pentode
1852	250	Triode

14.17. Signal and Noise in the Video System. The performance of an amplifier can be predicted from a knowledge of the fluctuation characteristics of the tube and coupling circuits. If the gain of the first stage of the amplifier in question is more than three or four, the only noise which produces any appreciable effect upon the signal-to-noise ratio will be that of the coupling resistance and that generated by the first tube. Where the

* See Kauzmann, reference 21.

gain of the stage is less than three, the noise of the first interstage coupling circuit and second tube must be taken into account.

The specific problem that will be considered in this section relates to the Iconoscope and its associated video amplifier, but the reasoning employed is applicable to any pickup system.

There are a number of ways in which the Iconoscope may be coupled to the amplifier. The simplest form of coupling consists of a resistance between the signal lead and ground. The shunt capacity of the Iconoscope and grid of the first tube is in parallel with this resistor. Unless the resistance of the resistor is quite low this capacity has an important effect on both the frequency response and the signal-to-noise ratio. In Chapter 10 the signal-to-noise ratio of an Iconoscope coupled with a 10,000-ohm resistance was estimated, leaving out of consideration the shunt capacity. However, since the frequencies involved are as high as 4 or 4½ megacycles, the effect of this capacity cannot be neglected, unless the coupling resistance is no greater than 2000 ohms.

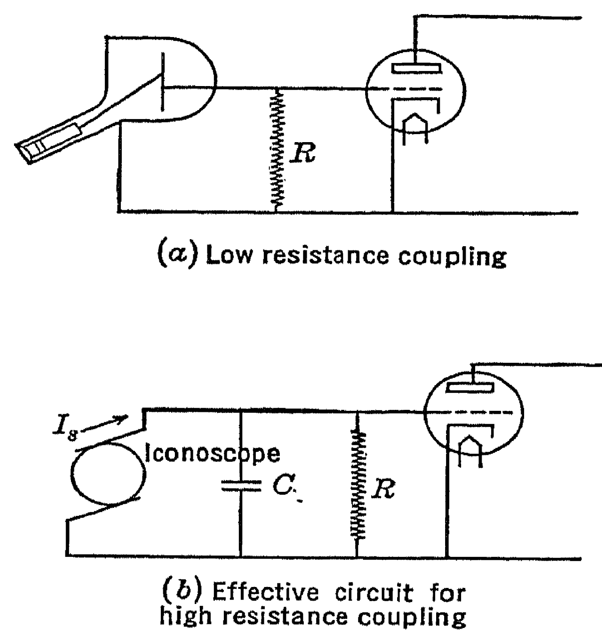


FIG. 14.33. — Iconoscope Coupling Circuits.

Where the coupling resistance is greater than 2000 ohms, the simple coupling shown in Fig. 14.33*a* should be represented by the effective circuit shown in Fig. 14.33*b*. The impedance of this circuit is

$$Z = \frac{R - 2\pi jfCR^2}{1 + 4\pi^2 f^2 C^2 R^2}. \tag{14.28}$$

Hence the magnitude of the signal impressed on the grid of the first tube is

$$V_i = |Z| i_s = \frac{Ri_s}{\sqrt{1 + 4\pi^2 f^2 C^2 R^2}},$$

where i_s is the signal current from the Iconoscope. In order to correct for the attenuation of the signal with increasing frequency, the amplitude response of the video amplifier must have the following form:

$$G = B\frac{\sqrt{1 + 4\pi^2 f^2 C^2 R^2}}{R}.$$

The signal voltage output is independent of frequency and is given by

$$V_s = Bi_s. \tag{14.29}$$

The resistive component which determines the noise generated by the coupling impedance is

$$\frac{R}{1 + 4\pi^2 f^2 C^2 R^2}.$$

This, together with the equivalent noise resistance of the first tube, R_t, determines the noise input to the amplifier. The input noise in the frequency range f to $f + df$ is

$$d\overline{V^2} = 4kT\left(\frac{R}{1 + 4\pi^2 f^2 C^2 R^2} + R_t\right)df.$$

In passing through the amplifier, the noise voltage is amplified in the normal manner. Therefore the mean square noise voltage appearing at the output of the amplifier over a frequency band extending from 0 to f is

$$\begin{aligned}\overline{V_N{}^2} &= \int_0^f 4kT\left(\frac{R}{1 + 4\pi^2 f^2 C^2 R^2} + R_t\right)\frac{B^2(1 + 4\pi^2 f^2 C^2 R^2)}{R^2}df \\ &= \frac{4kTB^2}{R^2}\left[(\mathrm{R} + R_t)f + \frac{4\pi^2 C^2 R^2 R_t f^3}{3}\right].\end{aligned} \tag{14.30}$$

The signal-to-noise ratio is obtained by dividing Eq. 14.29 by the rms noise voltage from Eq. 14.30:

$$S = \frac{V_s}{V_N} = \frac{Ri_s}{\sqrt{4kT}\sqrt{(R + R_t)f + \dfrac{4\pi^2}{3}C^2R^2R_tf^3}}. \tag{14.31}$$

If the coupling resistance R is small compared with the impedance of the shunt capacity, Eq. 14.31 reduces to:

$$S = \frac{Ri_s}{\sqrt{4kT(R + R_t)f}}, \tag{14.31a}$$

which is identical with the expression for the signal-to-noise ratio found in Chapter 10. On the other hand, when the resistor R is very large, the equation becomes:

$$S = \frac{1}{4\pi}\sqrt{\frac{3}{kT}}\frac{i_s}{C\sqrt{R_t f^3}}. \tag{14.31b}$$

It is interesting to note that the coupling resistance does not enter into the expression for the signal-to-noise ratio when this resistance is large.

For comparison, the two cases are evaluated for the following representative conditions:

Bandwidth	$f =$	4 megacycles
Shunt capacity	$C =$	15 micro-microfarads
Equivalent noise resistance	$R_t =$	1000 ohms

Case I. Eq. 14.31*a*

The resistance coupling of the pickup tube and amplifier must be less than the capacitative reactance. To meet this requirement R is assumed to be 2000 ohms. The signal-to-noise ratio is

$$S = \frac{2000\, i_s}{\sqrt{4kT \cdot 4.10^6(1000 + 2000)}} = 1.4 \cdot 10^8 i_s.$$

Case II

Where the coupling resistance is of the order of 200,000 ohms, the expression giving the signal-to-noise ratio is given by Eq. 14.31*b*. Under the conditions given above this has the value

$$S = \frac{1}{4\pi}\sqrt{\frac{3}{kT}}\,\frac{i_s}{15 \cdot 10^{-12}\sqrt{1000 \cdot 64 \cdot 10^{18}}} = 5.7 \cdot 10^8 i_s.$$

The coupling employing a high resistance and an amplifier corrected for the attenuation at high frequencies can be seen from these examples to give more than four times the signal-to-noise ratio. The practical difficulties of obtaining adequate correction in an amplifier so coupled, which have in the past been an obstacle preventing its adoption, have now been largely overcome.

14.18. Blanking and Signal Insertion. It has already been pointed out that the complete video signal must contain the information necessary to synchronize the scanning pattern at the receiver in addition to the picture signal itself.

The signal obtained directly from an Iconoscope (or other pickup device) does not contain all the required information. To the Iconoscope signal must be added not only the synchronizing impulse whose formation is discussed in the next chapter, but also information as to the average illumination falling on the scene, since the tube is essentially an alternating-current device. Furthermore, a disturbance is generated dur-

ing the return time of the scanning beam, which must be removed from the signal output of the tube.

The disturbing signal can be removed by rendering the amplifier inoperative during the return time of the beam. This process is commonly known as blanking the amplifier. Blanking may be accomplished by driving the grid of one stage of the amplifier below cutoff for the length of time it is desired that the amplifier be inoperative. As a consequence of biasing the blanking stage to cutoff, a pulse is generated. Since this pulse is unidirectional and positive, excess portions can be easily removed by arranging the bias of the succeeding stage in such a way that the tube is saturated during the blanking interval, or by the use of any one of the many other conventional limiters. A second way of blanking the amplifier is shown in Fig. 14.34. This makes use of an additional compensating

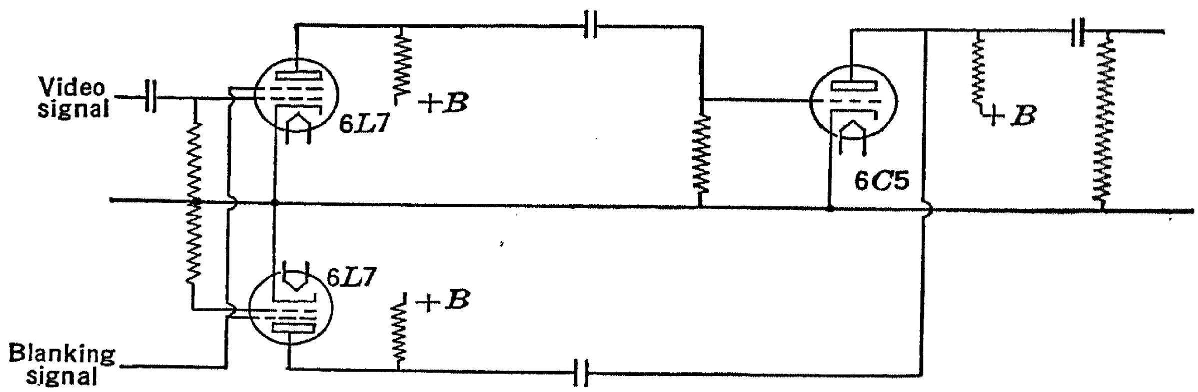


Fig. 14.34.—Balanced Blanking Circuit.

tube which balances out the effect of the blanking signal on the succeeding stages.

The addition of an impulse is an alternative to rendering the amplifier inoperative during the return time for blanking purposes. Thus an impulse of large amplitude whose duration is slightly in excess of return time may be added in the direction of a black signal to the video signal. The resulting composite signal is then put through a limiter which removes the top part of the impulse and, with it, the unwanted disturbance.

As was pointed out in Chapter 10, the Iconoscope does not yield information as to the average illumination of the scene. In other words, the signal generated by a gray line on a black background will be identical with that of a white line on a gray background. The signal representing either of these images before and after blanking is shown in Figs. 14.35*a* and *b*. To supply the missing information a phototube can be placed in the Iconoscope camera, so positioned that the light it receives is proportional to the average illumination in the picture. The output of this phototube, therefore, represents wanted information.

There are a number of ways of adding this information to the video signal. For example, the current from the phototube may be made to control an impulse generator in such a way that the amplitude of these impulses is proportional to the average light. This generator is so ad-

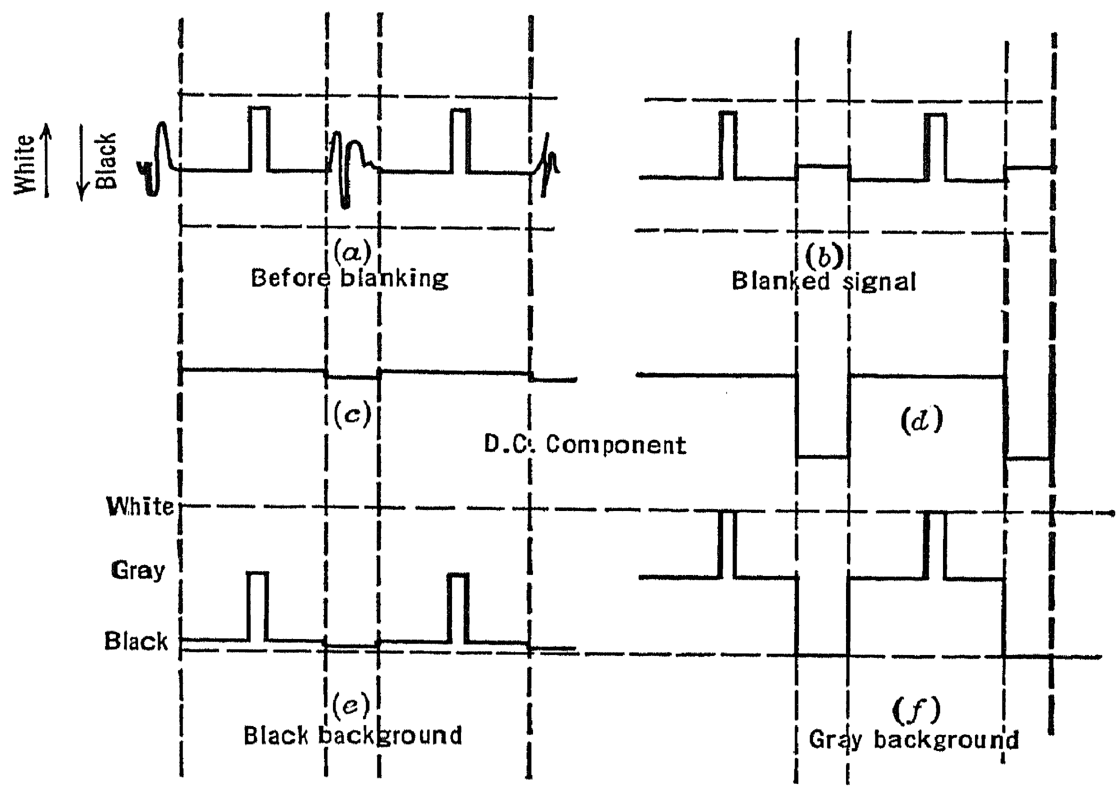


Fig. 14.35.—Formation of Video Signal with Pedestal Indicating Mean Light Level.

justed that the pulse it produces coincides both in time and duration to the blanking operation. The impulses from a gray bar on black and a white bar on gray are shown in Figs. 14.35*c* and *d*. These impulses are

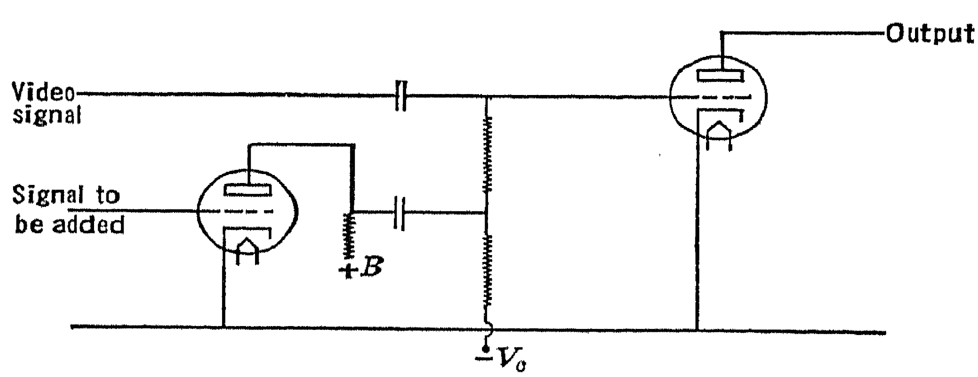


Fig. 14.36.—Simple Injector Circuit.

added directly onto the video signal, resulting in a final signal which contains the desired information shown in Figs. 14.35*e* and *f*. The average illumination may also be introduced by utilizing the amplified output of the phototube to control a blanking limiter.

The addition of an impulse such as that indicating the average illumination, or the synchronizing impulses, may be carried out merely by introducing the signal to be added at the bottom of the grid resistor of the stage where the addition is to take place. By inserting the signal in this way, additional shunt capacity is not introduced into the amplifier. A circuit of this type is shown in Fig. 14.36.

14.19. Video Cable. Often it is necessary to interconnect two remotely placed amplifiers. This requires a cable that can carry an extremely wide frequency band. A recently developed cable has proved to be very effective for this purpose. This cable consists of a small conductor in a relatively large metal shield. The conducting core is separated from the shield by means of small, low-loss insulators spaced at intervals of a few inches. In this way the capacity per unit length is made much

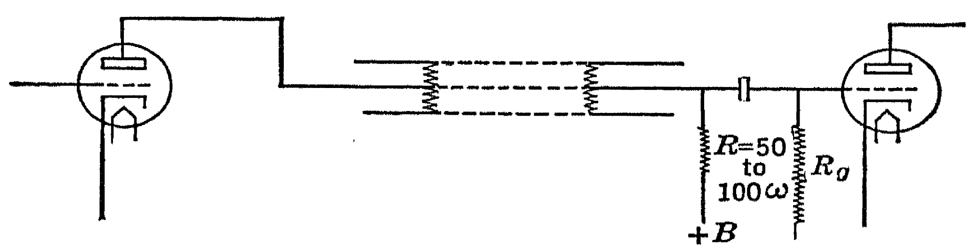


FIG. 14.37.—Plate-Coupled Driving Circuit for Cable.

smaller than it would be if the space between the shield and the conductor were filled with an insulating material. Furthermore, such a cable has a very low leakage conductance for a given length.

The capacity of this type of cable is between 15 and 25 micro-micro-farads per foot. Where short lengths are used, as for example between the pre-amplifier and the monitoring amplifier, the capacitative impedance is high enough so that, if the load at the output end has a fairly low impedance, the frequency and phase distortion produced by the cable is quite small and can be compensated by the peaking means described in earlier sections.

In order to couple the output of an amplifier efficiently to the cable the output tube must be capable of driving a low impedance. Because of the size of the coupling condenser that would be necessary with the conventional load resistance and capacitor, the circuit shown in Fig. 14.37 is commonly employed. The connection indicated requires that the core of the cable be at the plate potential of the driver tube.

The cathode coupled circuit shown in Fig. 14.38 is very well adapted to driving a low impedance. A tube coupled in this way cannot have more than unity voltage gain. However, because of the low impedance into which the output tube must work, the gain of this stage will, in general, be far below unity, irrespective of the circuit.

Where a long cable is necessary the problem of its termination is much more difficult. The termination must correct for the attenuation at high frequency and for the form of the phase delay. It must also match the impedance of the cable in order to avoid reflections. In other words, the

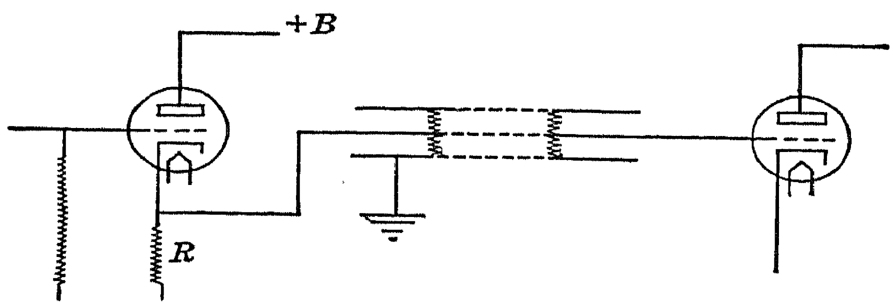


FIG. 14.38.—Cathode-Coupled Driving Circuit for Cable.

termination must equal the surge impedance so that the cable will act as though it were infinite in length.

The surge impedance of a cable is given by:

$$Z_s = \sqrt{\frac{r + j\omega L}{g + j\omega C}},$$

where r and g are the resistance and leakage conductance per unit length, L and C the inductance and capacity also per unit length. This becomes invariant with frequency if:

$$\frac{r}{g} = \frac{L}{C}.$$

However, it is not practicable to make a cable having this characteristic as the leakage conductance would have to be so high that the loss would be extremely great. Therefore, the input impedance of the terminus must vary with frequency so as to match the surge impedance, and then the attenuation of this unit must vary with frequency in order to produce constant overall response. This usually requires an attenuation which decreases with increasing frequency. The surge impedance of a television cable such as described is about 80 ohms and increases to a very high value at low frequencies.

14.20. The Complete Amplifier. In order to knit together the various elements discussed in this chapter, Fig. 14.39 is given, showing a circuit diagram of a typical pre-amplifier used in television work. The complete amplifier is shown as a block diagram in Fig. 14.40, together with the signal injection points.

This chapter no more than outlines the basic principles involved in video amplifier design. It is not intended that this summary should be

used in the actual design and construction of an amplifier, as a treatment of the details of this problem obviously requires a textbook in itself.

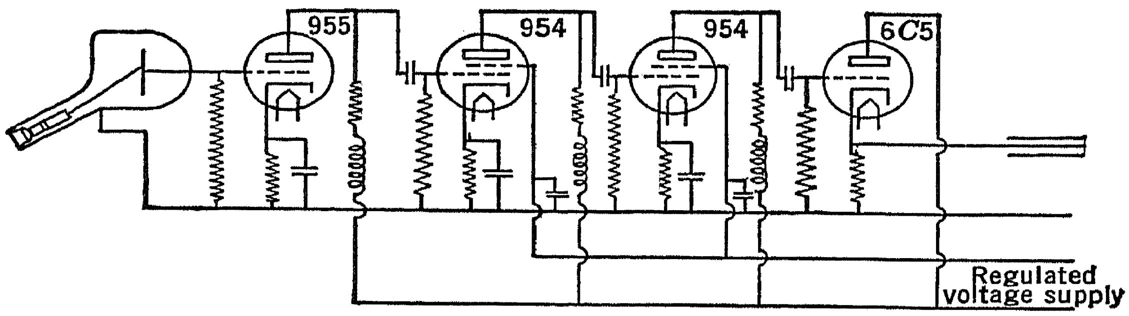


Fig. 14.39.—Typical Iconoscope Preamplifier.

Those interested are referred to the many original articles on this subject, some of which are listed in the bibliography immediately following.

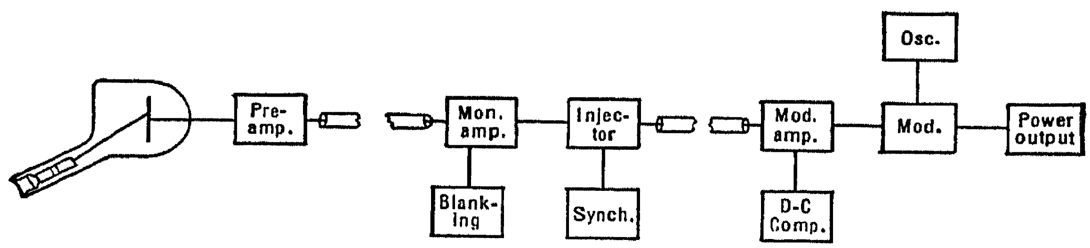


Fig. 14.40.—Block Diagram of Complete Amplifier Including Injector and Blanking Units.

REFERENCES

1. J. C. Wilson, "Television Engineering," Pitman, London, 1937.
2. F. E. Terman, "Radio Engineering," McGraw-Hill, New York, 1937.
3. K. McIlwain and J. G. Brainerd, "High Frequency Alternating Currents," John Wiley & Sons, New York, 1931.
4. T. E. Shea, "Transmission Networks and Wave Filters," D. Van Nostrand, New York, 1929.
5. E. L. Chaffee, "Theory of Thermionic Vacuum Tubes," McGraw-Hill, New York, 1933.
6. G. D. Robinson, "Theoretical Notes on Certain Features of Television Receiving Circuits," *Proc. I. R. E.*, Vol. 21, pp. 833–843, June, 1933.
7. S. W. Seeley and C. N. Kimball, "Analysis and Design of Video Amplifiers," *R. C. A. Rev.*, Vol. 2, pp. 171–183, October, 1937, and Vol. 3, pp. 290–308, January, 1939.
8. A. Preisman, "Some Notes on Video Amplifier Design," *R. C. A. Rev.*, Vol. 2, pp. 421–432, April, 1938.
9. F. A. Everest, "Wide-Band Television Amplifiers," *Electronics*, Vol. 11, pp. 16–19, January, 1938.
10. E. W. Herold, "High Frequency Correction in Resistance Coupled Amplifiers," *Communications*, Vol. 18, pp. 11–14, August, 1938.
11. H. A. Wheeler, "Wide-Band Amplifiers for Television," *Proc. I. R. E.*, Vol. 27, pp. 429–438, July, 1939.

12. A. V. Bedford and G. L. Fredendall, "Transient Response of Multistage Video-Frequency Amplifiers," *Proc. I. R. E.*, Vol. 27, pp. 277–284, April, 1939.
13. H. A. Wheeler, "The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes," *Proc. I. R. E.*, Vol. 27, pp. 359–384, June, 1939.
14. O. J. Zobel, "Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," *Bell System Tech. J.*, Vol. 7, pp. 438–534, July, 1928. See also "Electric Wave Filters," *Bell System Tech. J.*, Vol. 10, pp. 284–341, April, 1931.
15. J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Phys. Rev.*, Vol. 32, pp. 97–109, July, 1928.
16. H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, Vol. 32, pp. 110–113, July, 1928.
17. F. B. Llewellyn, "A Study of Noise in Vacuum Tubes and Attached Circuits," *Proc. I. R. E.*, Vol. 18, pp. 243–265, February, 1930.
18. G. L. Pearson, "Fluctuation Noise in Vacuum Tubes," *Physics*, Vol. 5, pp. 233–243, September, 1934.
19. B. J. Thompson and D. O. North, "Statistical Current Fluctuations in Vacuum Tubes" (in press).
20. W. S. Percival and W. L. Horwood, "Background Noise Produced by Valves and Circuits," *Wireless Engineer*, Vol. 15, pp. 128–137, March, 1938; pp. 202–207, April, 1938.
21. A. P. Kauzmann, "New Television Amplifier Receiving Tubes," *R. C. A. Rev.*, Vol. 3, pp. 271–289, January, 1939.
22. R. D. Kell, A. V. Bedford, and M. A. Trainer, "An Experimental Television System," *Proc. I. R. E.*, Vol. 22, pp. 1246–1265, November, 1934.

CHAPTER 15

SCANNING AND SYNCHRONIZATION

The complete deflection system for scanning consists of two pairs of deflecting coils or plates, two sawtooth generators, and two discharge impulse oscillators. One set of these elements operates at horizontal frequency, the other at vertical or frame frequency. Both transmitter and receiver are, of course, equipped with such systems, which must be exactly synchronized. This is accomplished by transmitting, along with the picture, synchronizing signals which, by suitable selector circuits, can be separated into vertical and horizontal synchronizing impulses. These impulses are used to actuate or control the respective deflection generators at the receiver. In practice not only does the synchronizing signal control the receiver scanning, but also the deflecting circuit of the Iconoscope, i.e., the deflecting system at the Iconoscope does not generate the synchronizing impulse but rather both Iconoscope and receiver deflection are controlled by an independent synchronization generator.

In the United States a ratio of frame frequency to scanning line frequency of 441, and a frame frequency of 30 pictures per second, have been adopted by the Radio Manufacturers Association. Since interlaced scanning is used, the frequency of the vertical deflection is 60 cycles per second, two vertical strokes constituting a single complete picture period. The frequency of 30 frames per second has been chosen because most of the electrical power used in the United States is generated at 60 cycles and, by making the frame frequency a simple submultiple of this, much trouble from "hum," that is, alternating-current ripple, can be avoided. For similar reasons the British Broadcasting Company has adopted a repetition frequency of 25 frames per second and a vertical deflection frequency of 50 cycles per second.

15.1. Requirements of Scanning. The theory of scanning in relation to the analysis and synthesis of the picture itself was taken up in some detail in Chapter 6. From this consideration it is quite evident that the success of the transmission of the picture depends upon an exact correspondence in the position of the scanning beam in the Iconoscope and in the Kinescope. In order to maintain this correspondence of position it is necessary to employ rather special synchronizing and scanning equip-

ment. This equipment and its operation form the subject of the present chapter.

The treatment of the problem divides itself naturally into three parts: first, the means for producing the motion of the spot in the receiving and transmitting tubes; second, the generator required to supply the power for producing the motion of the spot; and, finally, methods of synchronizing the transmitter and receiver. However, before taking up these topics, some of the principles of scanning will be reviewed in order better to formulate the problem as a whole.

In a system using the Iconoscope and Kinescope as its two terminal tubes, scanning consists of causing the electron beam to sweep over the image area (either mosaic or viewing screen) in a series of straight parallel lines. The pattern thus generated is illustrated in Fig. 15.1. These lines may be laid down in sequence, or may be interlaced, that is, laid down in the order 1, 3, 5, etc., 2, 4, 6, etc.

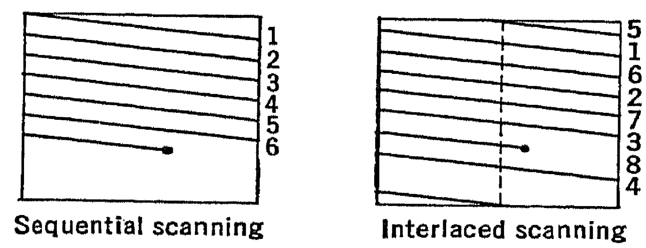


FIG. 15.1.—Sequential and Interlaced Scanning.

A perfect scanning pattern will be defined as one in which the following conditions are fulfilled:

1. The beam moves at a constant velocity along each line.
2. The lines are straight.
3. The lines are equally spaced.
4. The lengths of the lines are equal.

If the beam at the Iconoscope sweeps out a perfect pattern, then, in order that the received picture be a geometrically exact reproduction of the image being transmitted, it is necessary that the scanning pattern on the viewing tube or Kinescope be also perfect. Furthermore, the scanning beam at the transmitter and receiver must start and reach the end of the picture simultaneously. That these are sufficient conditions is obvious and requires no further explanation.

On the other hand, if the pattern at the Iconoscope is distorted, an exact reproduction of the image can be obtained at the Kinescope only if its scanning pattern is distorted in an identical manner. Fig. 15.2 illustrates an Iconoscope pattern whose shape is distorted owing to varying lengths of the scanning line, and shows the images obtained on a perfect and on a similarly distorted Kinescope pattern. As far as geometrical fidelity of the image is concerned, a television system with both the transmitting and receiving patterns distorted is perfectly possible. However, since a rectangular picture area is considered desirable, the use of

a distorted pattern does not permit the fullest use of the available picture area. Similarly, if the scanning pattern at the transmitting tube is not perfect on account of a varying velocity of the beam along the scanning line, a geometrical distortion of the received picture will occur at the Kinescope unless its scanning beam likewise moves with a non-uniform velocity. Similar non-uniformity in both patterns, while causing no geometrical distortion in the received image, is undesirable because it leads to unequal horizontal resolution in different parts of the picture. This is

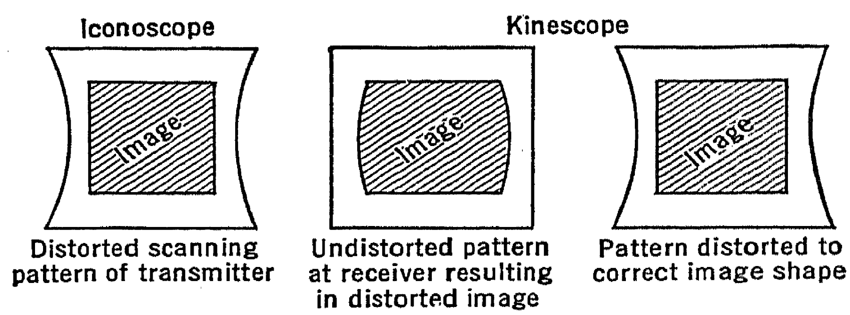


FIG. 15.2.—Effect of Distorted Pattern on Reproduced Image.

a consequence of the fact that, the higher the velocity of the scanning beam, the lower the resolution for a given band of transmitted frequencies. In general the most economical transmission of pictures requires as nearly perfect scanning patterns at the receiver and transmitter as are practicable.

Referring again to Fig. 15.1, it is apparent that the scanning pattern is generated by a displacement of the beam, which can be resolved into

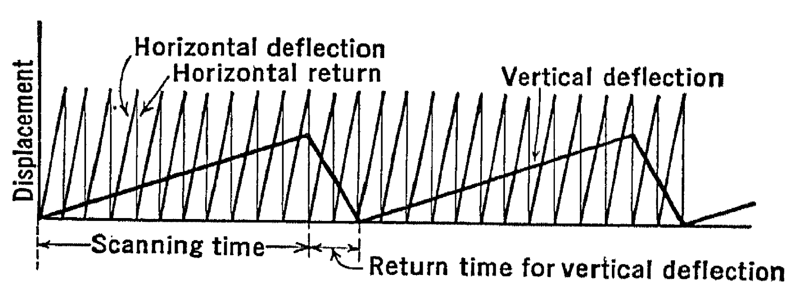


FIG. 15.3.—Sawtooth Displacement Components Required for Scanning.

two mutually perpendicular directions. Separating the motion of the scanning beam into its vertical and horizontal components, and plotting them as a function of time, sawtooth curves are obtained. The period of the sawtooth wave representing vertical displacement is the length of time required for the scanning beam to cover the entire picture area, or in other words, one frame or field period; the period of the horizontal sawtooth wave is the length of time required by the scanning beam to produce one line. The ratio of the vertical to the horizontal period is, of

course, equal to the number of lines making up the scanning pattern. Sawtooth waves which may represent the horizontal and the vertical displacements are shown in Fig. 15.3. It will be seen that each wave consists of a gradual motion of the beam across the image area and an abrupt return to its initial position. Since the only useful portion of the cycle is the gradual displacement, it is desirable to make the return time as short as possible. In practice it is found that the vertical return time cannot be made less than about 7 per cent of the total period, nor the horizontal return time less than 15 per cent. This return time is not entirely wasted as part of it is used for the transmission of the synchronizing impulse.

15.2. Deflection of Electron Beam. The cathode-ray beam from a properly centered electron gun in a Kinescope or Iconoscope will, if undisturbed, strike the surface of the fluorescent screen or mosaic near its center.* In order to produce scanning, the beam must be moved across the screen independently in two directions at right angles to each other.

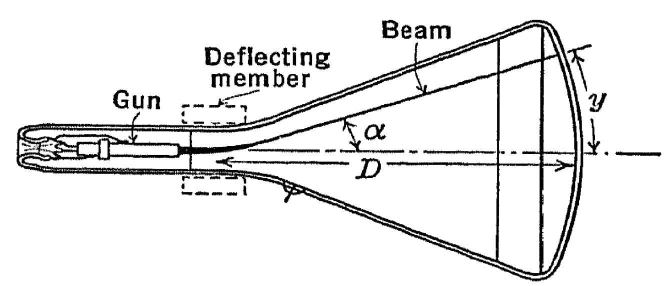


FIG. 15.4.—Deflection of Beam in Kinescope.

With a conventional gun, such as described in Chapter 13, one of the most satisfactory ways of displacing the spot is to deflect the beam through an angle close to the end of the gun, as is shown in Fig. 15.4. This deflection of the electron beam can be accomplished in either of two ways. An electrostatic field can be established in the region near the end of the gun and parallel to the direction in which the beam is to be deflected, or a magnetic field may be set up at right angles to the direction of deflection. Each of these methods has been successful in practical tubes as a means for producing horizontal and vertical deflection. Until quite recently the accepted practice was to use magnetic deflection in the vertical direction (frame frequency) and electrostatic deflection for the horizontal displacement (line frequency). One of the reasons for this choice was the difficulty in making an iron deflecting yoke which was sufficiently free from power loss at the high frequencies required for horizontal scanning. However, this objection to magnetic deflection is no longer of importance as the difficulties involved in the magnetic circuit have been completely overcome. The trend at present is toward the utilization of magnetic deflection for both vertical and horizontal displacement in nearly all large cathode-ray tubes. In certain special tubes,

* The angle of the mosaic relative to the axis of the gun positions the undeflected spot below the center to permit symmetric vertical angular displacement in scanning.

however, there are advantages in deflecting the beam electrostatically; therefore both methods have their role in television.

Electrostatic deflection is usually effected by means of a pair of parallel plates, placed at the end of the gun in such a way that the electron beam passes between them. The arrangement is shown in Fig. 15.5. In this figure G_1 and G_2 are the first and second anodes of the gun and $A - B$, $A' - B'$ are the two deflecting plates. A potential difference between the plates produces the electrostatic field which deflects the beam. Idealized field lines are shown between the plates. In order to calculate the deflection produced by these plates three simplifying assumptions will be made: first, that the length d of the plates is small compared with the distance between the center point of the plates and the fluorescent screen; second, that the plates are placed far enough from the first anode so that there is no interaction between the field of the second lens of the gun and the deflecting field; and finally, that the fringe fields have a negligible effect on the deflection. The error introduced by these assumptions will be considered later.

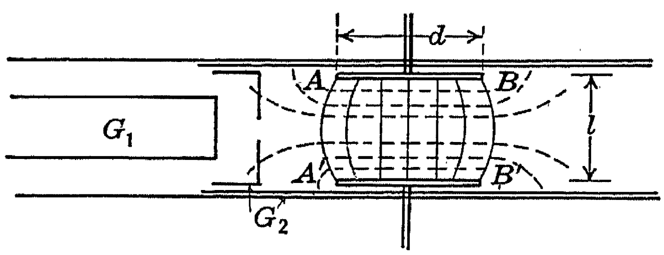


FIG. 15.5.—Arrangement of Deflecting Plates.

The field E between the plates is given by

$$E = \frac{V_A - V_{A'}}{l},$$

where V_A and $V_{A'}$ are the potentials applied to the deflecting plates and l is the separation between the plates. This field acts on the electrons leaving the gun over a portion of their path equal to the distance d. The electron path over this length is parabolic, and the displacement Δ from the axis is:

$$\Delta = \frac{1}{4}\left(\frac{V_A - V_{A'}}{V}\right)\frac{d^2}{l},$$

where the axial velocity in electron volts V is $(V_A + V_{A'})/2$. The ray is therefore deflected through an angle α, where:

$$\alpha \cong \tan \alpha = \frac{2\Delta}{d} = \frac{1}{2}\left(\frac{V_A - V_{A'}}{V}\right)\frac{d}{l}.$$

This deflection produces a displacement y of the spot at the screen which,

if D is taken as the distance between the screen and the center of the deflecting plates, is given by

$$y = \Delta + \alpha\left(D - \frac{d}{2}\right) = \frac{1}{2}\left(\frac{V_A - V_{A'}}{V}\right)\frac{d}{l}\left(\frac{d}{2} + D - \frac{d}{2}\right)$$
$$= \frac{1}{2}\left(\frac{V_A - V_{A'}}{V}\right)\frac{dD}{l}. \qquad (15.1)$$

This equation indicates that within the limits of the approximation the displacement on the screen is proportional to the potential difference between the deflecting plates. Therefore a sawtooth voltage wave must be applied to the deflecting plates to produce the displacement of the beam required for the scanning.

Except for very short tubes, the error made in assuming that the length of the deflecting plates d is small compared with the screen to deflecting plate distance D is, for all practical purposes, negligible. With short tubes, for example certain types of projection tubes, this approximation cannot be made, and the linear relation between the deflecting plate voltage and the displacement does not apply. In order to obtain sawtooth deflection of the beam in this type of tube either the deflecting plates must be specially shaped or the voltage wave form not strictly a sawtooth.

The second assumption is very closely fulfilled in practice as it is a necessary condition for the avoidance of the aberration of the spot which would otherwise be caused by the cylindrical lens formed by the deflecting plates and the gun elements. Furthermore, in order to minimize distortion of the lens fields of the gun by the deflection voltages the potentials applied to the deflecting plates usually are such that the second anode potential V_2 lies midway between them. In other words, $V_A - V_2$ is equal to $V_2 - V_{A'}$.

In practice the fringe fields around the deflecting plates cannot be neglected. These fringe fields have three major effects: first, they introduce a cylindrical lens into the electron-optical system of the gun, which may change its focusing properties so that the position of minimum spot size is made a function of the deflection, and even more important, may distort the shape of the spot; second, they make the displacement of the spot on the screen a function, not only of the potential applied to the deflecting plates, but also of any displacement of the beam in the plane of the plates, introducing "cross talk" between horizontal and vertical deflection; third, the fringe field may result in a slight departure from linearity in the relation between deflecting potential and displacement.

In order to overcome these effects it is necessary to use specially shaped or positioned deflecting plates rather than plane, parallel plates.

As has already been pointed out, it is necessary to deflect the beam in two mutually perpendicular directions. This can be done by having two pairs of deflecting plates, one ahead of the other, so that between the first the beam is deflected through an angle in one plane and then bent through an angle perpendicular to this between the second pair of plates.

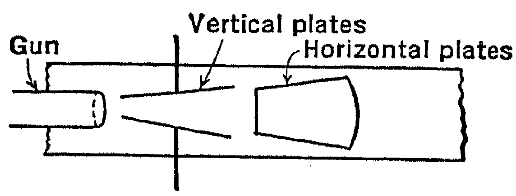


FIG. 15.6.—Successive Electrostatic Deflection in a Vertical and Horizontal Direction.

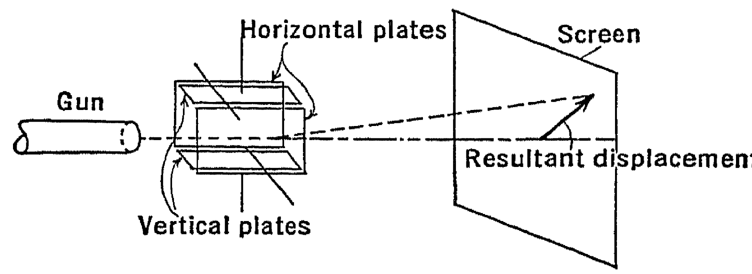


FIG. 15.7.—Combined Vertical and Horizontal Electrostatic Deflection.

A second method is to establish between two sets of plates a single field which deflects electrons in a direction corresponding to the resultant of the two components of displacement. These two deflection arrangements are illustrated in Figs. 15.6 and 15.7. While the second method is the more rational of the two, it is found to be much more difficult from the standpoint of properly shaping the plates to avoid spot and pattern distortion. Except in special experimental devices, all electrostatically deflected tubes at present use the first described arrangement.

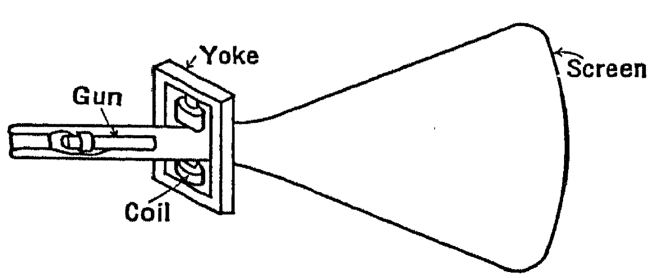


FIG. 15.8.—Simple Iron-Cored Coils for Magnetic Deflection of Beam.

Like electrostatic deflection, magnetic deflection is accomplished by establishing a field close to the end of the gun. The deflection is, of course, at right angles to the direction of the magnetic field lines. The field is produced by a pair of iron- or air-cored electromagnets placed outside of the gun neck. A pair of iron-cored deflecting coils is shown schematically in Fig. 15.8.

With the aid of certain simplifying assumptions, the displacement of the electron spot in terms of the magnetic field can be readily calculated. The first approximation is that the magnetic field strength H is uniform over a region of length d, and zero outside of this region. This simplification entirely neglects fringe field effect. The second assumption is that

the distance D from the center of the field to the screen is large compared with the region of field d, and to the displacement y on the screen.

From Fig. 15.9 it is seen that in the region occupied by the field the electrons move along a circular path and emerge traveling in a direction which makes a small angle α with their original motion. In the field-free region beyond the deflecting field the paths are straight. The electron stream arrives at the screen at a point displaced an amount y from its undeflected position.

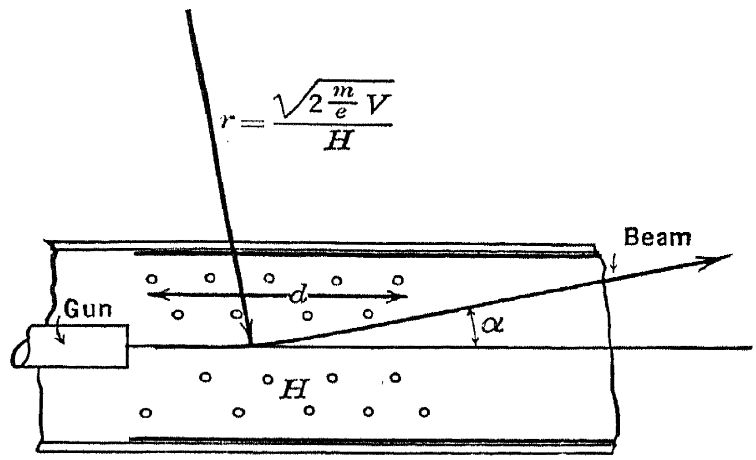


FIG. 15.9.—Deflection of Beam in Magnetic Field.

The force acting on an electron at right angles to its direction of motion is:

$$F = eH\sqrt{2\frac{eV}{m}}.$$

Therefore it will move in a circular path of radius:

$$R = \frac{\sqrt{2\frac{m}{e}V}}{H}.$$

The angle α through which it is deflected in traversing the distance d is therefore:

$$\alpha \cong \sin \alpha = \frac{d}{R} = \frac{Hd}{\sqrt{2\frac{m}{e}V}},$$

and the displacement Δ at the end of the circular path:

$$\Delta = R^2 - \sqrt{R^2 - d^2} \cong \frac{d^2}{2R} = \frac{d^2H}{2\sqrt{2\frac{m}{e}V}}.$$

The total displacement y is therefore:

$$y = \Delta + \alpha\left(D - \frac{d}{2}\right).$$

This may be written as:

$$y = \frac{d^2H}{2\sqrt{2\frac{m}{e}V}} + \frac{dH\left(D - \frac{d}{2}\right)}{\sqrt{2\frac{m}{e}V}}$$

or

$$y = \frac{DdH}{\sqrt{2\frac{m}{e}V}}. \tag{15.2}$$

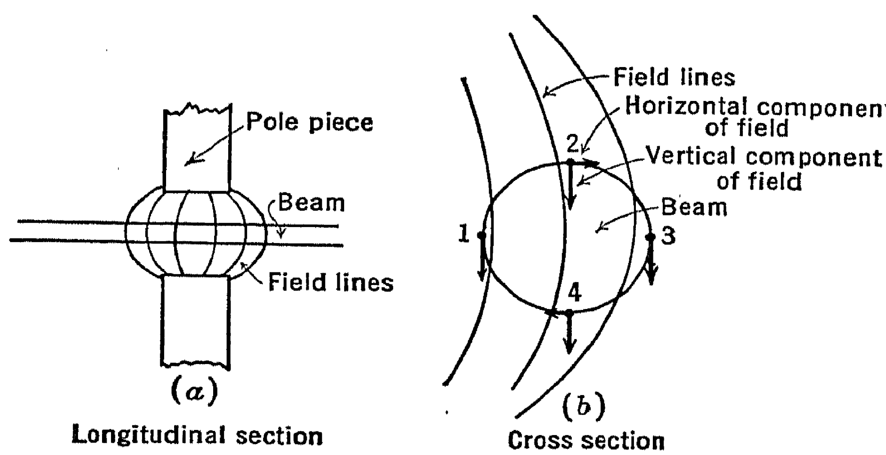


FIG. 15.10.—Influence of Non-Uniform Field on Electron Beam.

It will be noticed that the deflection for a given field H is inversely proportional to the square root of the second anode voltage, whereas in the case of electrostatic deflection it is inversely proportional to the voltage itself. This is an important factor in favor of magnetic deflection where high beam voltages are required, as, for example, in projection tubes.

Just as for electrostatic deflection, it is not possible to have a sharply defined region of uniform field. This can readily be seen from the form of the Laplace equations governing the magnetic field in free space. The fringing of the field tends to produce aberrations in the electron beam and also to distort the scanning pattern.

An exact analysis of the defects encountered in magnetic deflection is difficult in the extreme and can be carried out only with reference to a specified pair of coils and gun. However, the origin and nature of some of the difficulties can be shown rather easily.

Fig. 15.10 illustrates an electron beam under the influence of a non-uniform magnetic deflecting field. The lack of uniformity, both in diagram (*a*) of the plane including the axis of the beam and the pole pieces, and diagram (*b*) which shows a cross-section of the beam, has been greatly exaggerated for the sake of clarity. Since the cross-section of the beam has finite size the field is not constant over its area, causing different parts of the beam to be deflected different amounts. Along the diameter from point (1) to point (3) in diagram (*b*) the field decreases; therefore electrons in the neighborhood of (1) are deflected more than those near (3), resulting in a compression of the spot along this diameter. Furthermore, since the field is not uniform, the magnetic field lines are curved as shown. As a consequence of this curvature, the direction of the field at the top and bottom of the beam is not the same. Considering points (2) and (4), there are, as well as the field components producing the desired deflection, components parallel to the direction of deflection. This causes the electrons in the top portion of the beam to be bent upward, and those in the lower section downward. The net effect, therefore, is that the beam is compressed in the plane of deflection and expanded at right angles to it, resulting in an elliptical spot. If the beam is not in the median plane between the two poles, the effect is more complicated, but also leads to an elliptic spot whose major and minor axes may be rotated at arbitrary angles with respect to the direction of deflection.

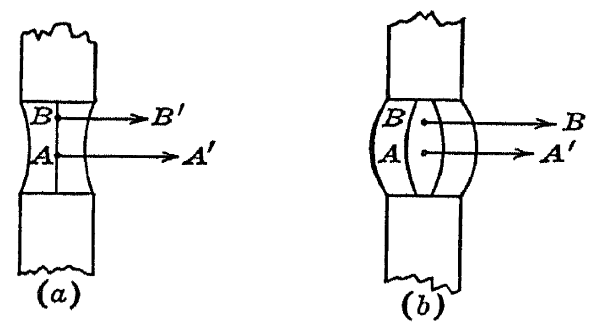


FIG. 15.11.—Cause of Scanning Pattern Distortion by Non-Uniform Field.

That the non-uniform field distorts the scanning pattern can be readily seen from Fig. 15.11, which again shows a section of the horizontal deflecting field normal to the beam. If the beam is undeflected in the vertical direction so that it lies in the median plane between the horizontal pole pieces corresponding to position (*A*) of the diagrams, it will be deflected in this plane by an amount represented by the vectors AA'. For convenience it may be assumed that this deflection is equal for the two types of distorted fields illustrated. However, if the beam has been deflected in the vertical direction so that it passes through points (*B*), the horizontal deflections will no longer be equal, for in the field represented in (*a*) it has moved into a region of lower field strength, while in (*b*) it is in a region of higher field strength. Therefore if the field has "barrel"-shaped distortion (*b*), the resulting scanning pattern will be pincushioned,

while if the field lines are "hourglass" shaped (*a*) the pattern will have barrel distortion.

These defects can be minimized by shaping the pole pieces properly and distributing the windings so that a nearly uniform field pattern is produced.

The two components of deflection at right angles to each other can be obtained by causing the beam to traverse in succession two mutually perpendicular fields, or by means of a single field. Actually, in the case of magnetic deflection the latter is almost always used. The field required can be obtained by superimposing the fields from two pairs of coils at right angles to each other. However, to produce a field which gives a

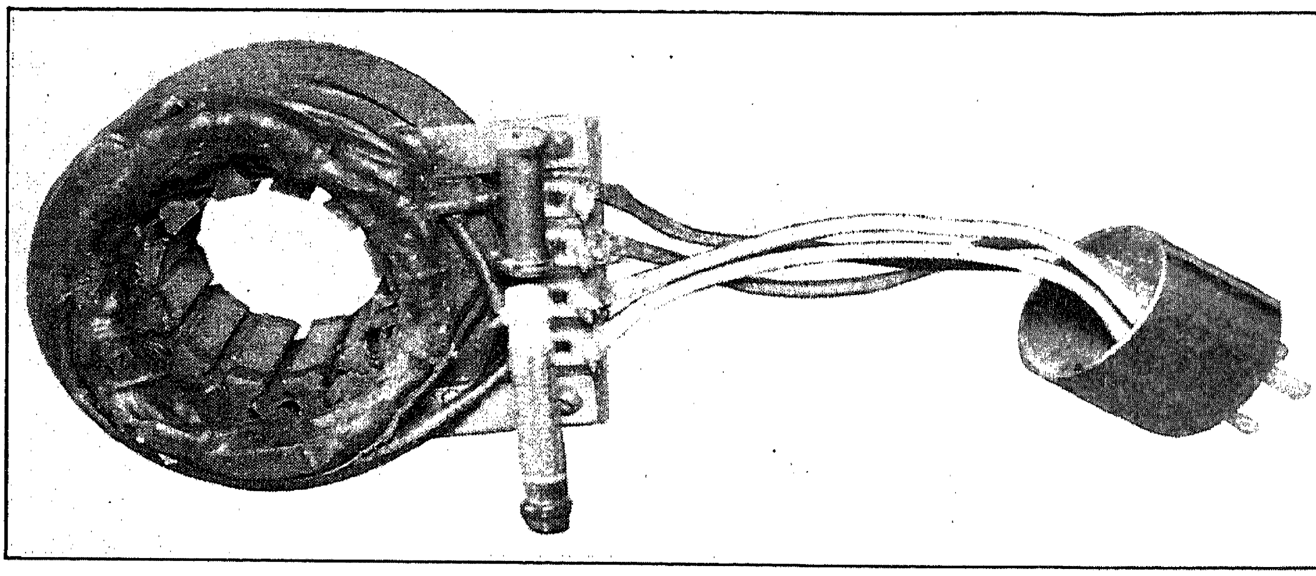


FIG. 15.12.—Iron-Core Deflecting Yoke Designed to Minimize Spot and Pattern Distortion (after Maloff and Epstein, "Electron Optics in Television").

scanning pattern free from distortion and a minimum defocusing of the spot requires very specially designed coils and cores. An example of a deflecting unit designed to minimize spot and scanning pattern defects having an iron core and distributed windings is illustrated in Fig. 15.12.

Although for most purposes an iron-cored unit is superior, occasionally air-cored deflecting coils are advantageous. Where iron is not present there can be no variation in permeability, therefore the reluctance of the magnetic circuit is constant, and the flux density at any point is at all times proportional to the deflecting current. Furthermore, magnetic saturation is eliminated by the use of air cores. However, the number of ampere-turns required to produce a given deflecting field is greater. This does not necessarily mean that more power is required to produce a given deflection, either at high or low frequency, but it does mean that this type of deflecting coil is more difficult to match to the usual deflection

generator. Also, this system lacks some of the flexibility of the former because there are no iron pole pieces which can be shaped to aid in obtaining a uniform field. A typical set of air-core deflecting coils is shown in Fig. 15.13.

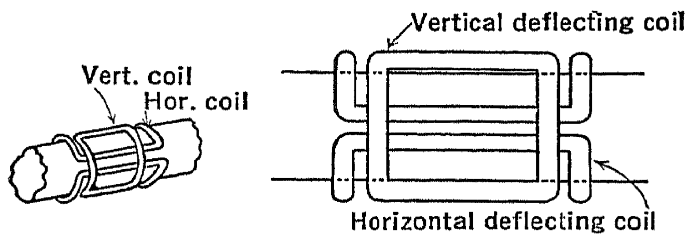


Fig. 15.13.—Simple Form of Air-Core Deflecting Coils.

15.3. The Deflection Generator. Assuming suitably shaped coils and cores, or correctly designed electrostatic plates, a perfect scanning pattern can be obtained only if they are supplied with current or voltage having a very specific time variation.

The wave form of the current through the coils producing the vertical deflecting field of the magnetic system must be sawtoothed and have a periodicity equal to the field frequency of the pattern. That supplied to the horizontal coils must also be sawtoothed, but its periodicity is much higher, corresponding to line frequency. The current in each case builds up linearly with time for most of the cycle, and then returns abruptly to its initial value. It is obviously impossible for the return to be instantaneous, because this would necessitate an infinite voltage across the coils owing to their finite inductance. Therefore a certain amount of time must be allowed to re-establish initial conditions, which interval is known as the return time. As has already been pointed out, this return time in practice occupies about 7 per cent of the vertical deflection period and about 15 per cent of the horizontal. For electrostatic deflection it is the voltage wave rather than the current which must have the sawtooth form.

The generator which supplies the power to these deflecting elements consists in general of an oscillator, a sawtooth generator, and, often, a deflection amplifier. The oscillator when in use is not free running but is set off by a series of synchronizing impulses which serve to coordinate the timing of the patterns at the transmitter and receiver. The sawtooth generator may include circuit elements in common with the oscillator in the simpler deflection units, but in general they should be considered as separate entities. It is frequently necessary to amplify the output of the sawtooth generator in order to obtain a current or voltage wave of sufficient amplitude to deflect the beam.

In order to better formulate the circuit requirements of the deflecting

system, particularly where the current or voltage wave must be amplified, an analysis of the harmonic content of the sawtooth wave is re-

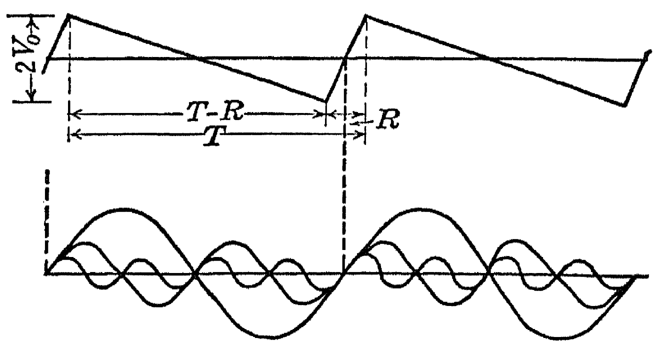


FIG. 15.14.—Sawtooth Wave and Its Fourier Components.

quired. The wave form, which is shown in Fig. 15.14, can be represented by a Fourier series as follows:

$$V(t) = b_0 + b_1 \sin \frac{2\pi}{T}t + b_2 \sin \frac{4\pi t}{T} + \cdot \cdot \cdot \; b_n \sin \frac{2\pi nt}{T} \qquad (15.3)$$

$$b_n = \frac{2}{T}\int_0^T V(t) \sin \frac{2\pi nt}{T} dt,$$

where

$$V(t) = \frac{2V_0 t}{R} \qquad -\frac{R}{2}<t<\frac{R}{2}$$

$$= \frac{V_0(T - 2t)}{T - R} \qquad \frac{R}{2}<t<T - \frac{R}{2}.$$

Carrying out the indicated integration, the coefficients for the series are:

$$b_n = V_0 \frac{2T^2}{(\pi n)^2 R(T - R)} \sin \frac{\pi nR}{T}. \qquad (15.3a)$$

The amplitudes of the first ten harmonic components for a 15 per cent return time ($R = 0.15T$), are given in Table 15.1. It will be seen from this table, and from the form of Eq. 15.3a, that the coefficients go through zero at frequencies for which $n = T/R$. Usually in designing the amplifiers and circuits required to handle the sawtooth voltage or current waves it is not necessary to consider frequencies beyond the point where the coefficients go through zero, unless the return time is extremely long. The coefficient b_0, representing the constant term of the series, is zero for a centered pattern with magnetic deflection, and is equal to the second anode voltage for electrostatic deflection (where the sawtooth represents the voltage on either plate relative to cathode potential).

TABLE 15.1

Amplitude of Harmonics in Sawtooth Wave

n	b_n/V_0
1	0.72154
2	0.32146
3	0.17442
4	0.09448
5	0.04496
6	0.01364
7	−0.00508
8	−0.01460
9	−0.01748
10	−0.01590

Assuming that ten harmonics are required for the sawtooth wave, the amplifier and circuits for horizontal deflection must respond to frequencies from 13,000 to 130,000 cycles per second; those for vertical deflection, from 60 to 600 cycles.

15.4. The Sawtooth Generator. The early forms of cathode-ray deflection systems employed a combined oscillator and sawtooth generator. A generator illustrative of this type is shown in Fig. 15.15. The circuit consists of a condenser, shunted by a two-electrode gas-filled tube, and supplied from a high-voltage source through the large resistor R. The condenser is charged through the resistor, and the voltage across it increases nearly at the constant rate $V(t) = V_B t/(CR)$ as long as $V(t)$ is small compared to V_B. When the potential of the condenser reaches the value V_0 (still small compared with V_B) the voltage is enough to cause the gas tube to break down, discharging the condenser. The resistor is so proportioned that the discharge in the gas tube stops when the condenser is discharged. The cycle then starts over again. The frequency of this cycle is determined by V_B, C, R, and the tube characteristics. It is adjusted to be slightly lower than the scanning frequency required. A small voltage impulse, if applied when the condenser is nearly charged to the breakdown voltage of the gas tube, will start the discharge. Impulses of this type can be used as a means of synchronizing the sawtooth with the required scanning frequency.

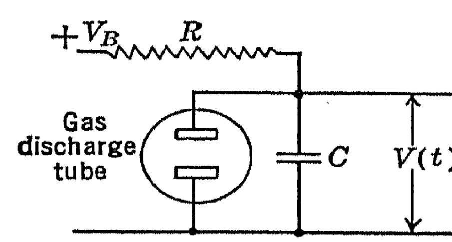


Fig. 15.15.—Simple Sawtooth Generator with Two-Element Gas Discharge Tube.

The sawtooth generator in this unit consists of the condenser C charged through R and discharged by the gas discharge tube. The oscillator is this same discharge tube, caused to break down periodically by the voltage built up across the condenser. The stability of this circuit as an oscillator depends upon the characteristic of the gas discharge tube. As these characteristics are in general not invariant throughout the life of the tube, a large controlling impulse is required to insure synchronization.

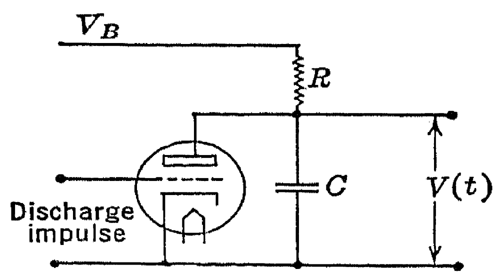


FIG. 15.16.—Grid-Glow Tube Sawtooth Generator.

To avoid the difficulties consequent on the requirement of a synchronizing impulse of large amplitude, a grid-controlled discharge tube may be substituted for the diode.

A circuit of this type is shown in Fig. 15.16, using a grid-glow tube such as a neon thyratron as the discharge tube. When the grid is driven positive the tube becomes conducting, and remains so until the condenser is fully discharged, irrespective of whether or not the grid is driven negative again before the discharge is complete. Obviously, then, the length of the return time depends upon the conductivity of the discharge tube, and the capacity and charge in the condenser, rather than upon the shape of the impulse from the oscillator.

Because of the difficulty of maintaining a perfectly constant return time, and because the quenching of the discharge tends to be too slow

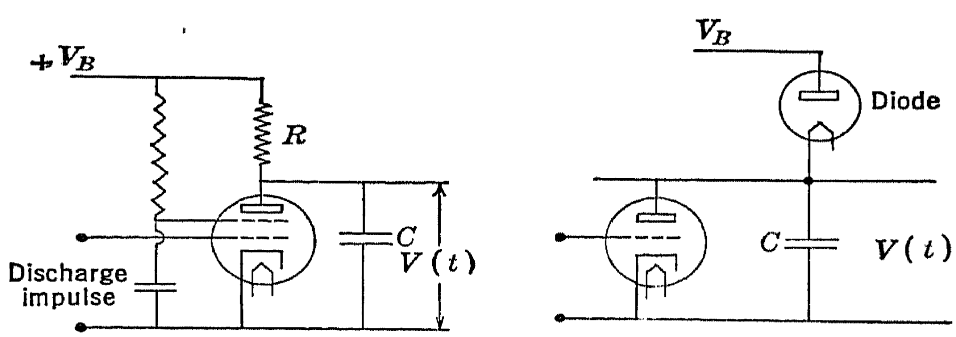


FIG. 15.17.—Sawtooth Generators Employing Hard Vacuum Tubes.

for the very high deflection frequencies needed in present-day high-definition television, this type of tube is not favored in the sawtooth generator of commercial receivers manufactured in this country.

A thermionic vacuum tube employed as the discharge element overcomes these difficulties. The duration of the discharge is accurately determined by the length of time the grid is maintained more positive than cutoff by the control impulse, and the tube cuts off in a negligibly short

length of time when the grid is swung negative. The circuits of sawtooth generators using this type of tube are diagrammed in Fig. 15.17.

In order to obtain a sawtooth wave which has a linear time variation each cycle, the voltage $V(t)$ across the condenser must be small compared with V_B. This can readily be seen from the expression giving the charge on the condenser, which is:

$$V(t) = V_B(1 - e^{-\frac{t}{CR}}),$$

and which can be expanded into the series

$$\frac{V(t)}{V_B} = \frac{t}{CR} - \frac{1}{2!}\left(\frac{t}{CR}\right)^2 + \frac{1}{3!}\left(\frac{t}{CR}\right)^3 - \cdot \cdot \cdot \cdot$$

For values of t/CR very small compared with unity, terms containing powers of the variable which are greater than one may be neglected. This condition is fulfilled only if $(V(t)/V_B) \ll 1$. As a consequence, the sawtooth

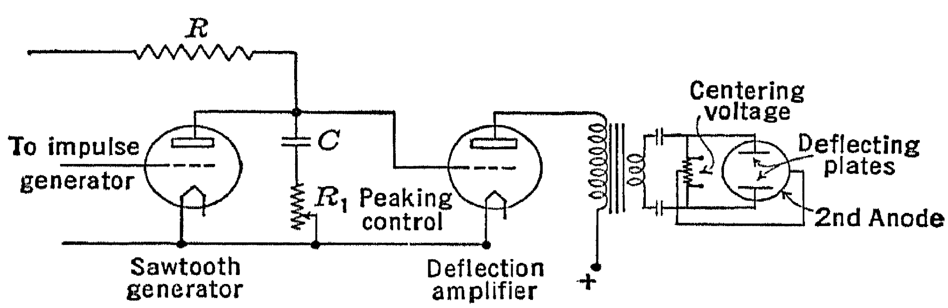


Fig. 15.18.—Sawtooth Generator and Amplifier.

generator is of necessity very inefficient. The inefficiency can be overcome if a temperature-limited diode or any other constant current device is substituted for the resistance R. In practice, however, this is rarely necessary since economy in the generator is relatively unimportant.

Ordinarily the amplitude of the sawtooth wave produced by such a generator is not sufficiently great to apply directly to the deflecting plates without amplification. Of course, it is possible to obtain any voltage desired by applying a sufficiently large voltage to the charging resistor. However, since a deflection voltage of 1000 or more volts is often required, an overall voltage of several thousand volts would be necessary. Such voltages would greatly complicate the practical design and construction of the circuit.

The output from the generator is consequently fed into an amplifier which supplies the deflecting plates. When a transformer is used it tends to distort the wave form on account of the leakage reactance of the windings. To correct for this it is necessary to add to the sawtooth voltage wave a rectangular wave. A resistance R_1 in series with the charging con-

denser C adds the required rectangular wave. By varying this resistance the amount of square wave can be varied and the shape of the deflection voltage controlled. Coupling the deflecting plates to the deflection amplifier with a transformer makes the problem of swinging the two deflection plates symmetrically about the second anode voltage very simple, as a center tap in the secondary of the transformer can be tied to the second anode potential.

With magnetic deflection, the current through the coils must have a sawtooth wave shape. If the coils are considered as perfect inductances, the voltage wave required to give a sawtooth current would be a rectan-

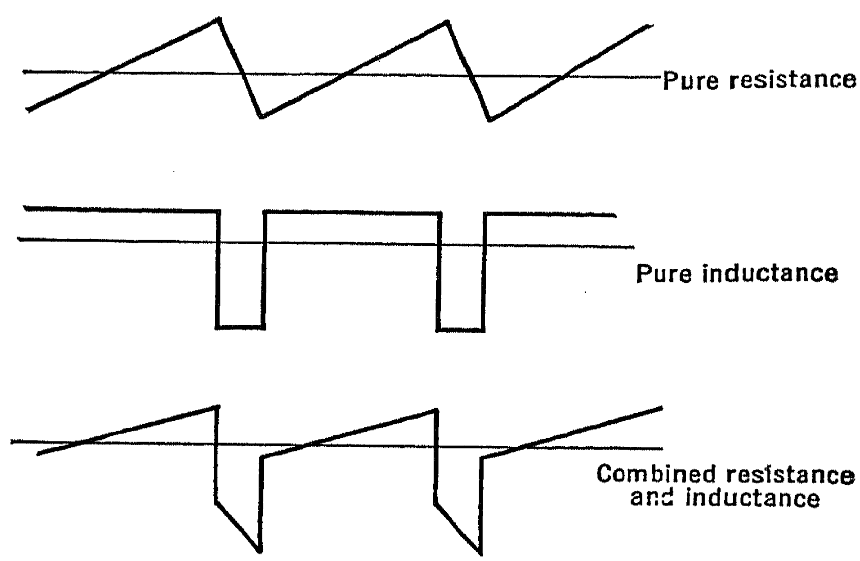


FIG. 15.19.—Voltage Wave Shapes Required for Sawtooth Deflection.

gular impulse wave. The actual circuit consists of the plate resistance of the driving amplifier in series with the coils which themselves are not resistanceless. The voltage which must be applied to the grid of the driving stage is therefore given by the equation

$$e_1(t) = \frac{1}{\mu}\left(ir + L\frac{di}{dt}\right),$$

where L is the inductance of the coils and r the total circuit resistance. From this relation it will be seen that over the sweep period $0 < t < T - R$ the voltage, except for a constant component, is:

$$e_1(t) = \frac{I_0}{\mu(T - R)}(rt + L),$$

while during the return time $T - R < t < T$, it is:

$$e_1(t) = \frac{I_0}{\mu R}(Tr - tr - L).$$

The voltage wave shape therefore depends upon the relative values of r and L. It is shown for three different conditions in Fig. 15.19.

The rectangular wave shape required for a circuit which is primarily inductive is obtained across a resistance in series with the condenser of the sawtooth generator. A combined sawtooth and rectangular voltage wave or a sawtooth wave can be made available by methods already discussed.

The shape of the driving wave needed depends largely upon the type of output tube employed in the circuit. If this tube is a triode its plate resistance is low and the driving wave will be a combined rectangular and sawtooth wave. On the other hand, a pentode in the final stage makes the resistance of the circuit very high, and the required voltage is sawtooth in form.

A pentode stage, though efficient, introduces another difficulty because its high resistance offers little damping to any oscillations which

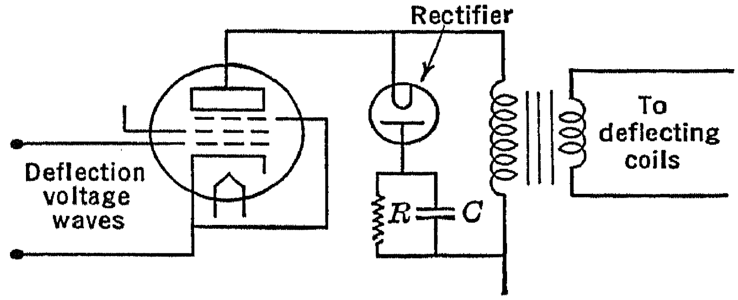


Fig. 15.20.—Diode Used in Special Damping Circuit.

may occur owing to the inductance and self-capacity of the coils. The abrupt voltage changes required for deflection tend to excite such oscillations. An ordinary RC damping circuit across the coils is usually unsatisfactory, as it prevents building up the large voltage required to bring about the rapid reversal of current during the return time. In order to obtain the damping without increasing the return time, a unidirectional damping circuit can be employed. This circuit consists of a diode in series with a parallel RC circuit as shown in Fig. 15.20. Under certain operating conditions such a unidirectional filter circuit can be made to increase the efficiency of deflection.

A transformer is ordinarily employed to couple the output stage to the deflecting coils. This is done to improve the impedance match between the normal deflecting unit and the driving tube. However, the introduction of a transformer does not alter the operating conditions discussed in the preceding paragraph.

Before leaving this subject it should be pointed out that the voltage across the deflecting coils and transformer during return time may be very large. This voltage is greater than the deflection voltage by a factor

equal to the ratio of deflection time to return time. Special care, therefore, must be exercised in designing the insulation of the deflecting yoke, transformers, etc.

15.5. The Control Oscillator. The sawtooth generators described in the preceding section require an impulse to trigger the discharge tube. Except for the generator circuit employing a simple gas-filled diode, this impulse not only maintains synchronization but also actuates the circuit. The control impulse which performs this function is a positive pulse which occurs at the end of the deflection and whose duration is equal to the return time.

The synchronizing impulses supplied to the deflection generator from the separator circuits which are described in section 15.7 are usually of insufficient amplitude to actuate the discharge tube of the sawtooth generator. Before they can be used for this purpose they must be amplified. This amplification could be effected by conventional methods with several cascaded amplifier stages. An alternative is to control a special oscillator by the separated synchronizing signal. The output of this oscillator is in the form of impulses of the required amplitude and shape to trigger the discharge tube.

The advantages of the latter method are twofold. In the first place, fewer tubes and circuit elements are needed to obtain the required discharge impulse, and in the second place the impulse continues in the absence of synchronization.

An ordinary resonant circuit sine-wave oscillator is not suited for this purpose, because elaborate shaping circuits are needed to give the required wave form. An even more serious objection is that the phase relation between the controlling impulses and the output depends upon the difference in frequency between the free oscillation of the tuned circuit oscillator and the impulses. This property results in a shift of the picture as a whole with any change in natural period or synchronizing frequency. The oscillator must have more nearly the characteristics of a flywheel than the pendulumlike attributes of a resonant oscillator, if this condition is to be avoided.

A relaxation oscillator, such as a dynatron oscillator or a multivibrator, can be used to produce the required impulses. The action of this type of oscillator is characterized by a period of violent change during which energy is stored up in a reactive element and a period of relative quiet during which this energy is dissipated or transferred. As a consequence of its mode of operation its output can easily be made to consist of impulses occurring at the periods of change.

A circuit for a dynatron relaxation oscillator is shown in Fig. 15.21. Relaxation oscillations are conditioned on a very high ratio of the in-

ductance to the shunt capacity C. When these circuit requirements are fulfilled, oscillations, whose wave form is shown in Fig. 15.22, are obtained. The condenser C_2 in the output of the oscillator, together with the grid resistor of the discharge tube, filters out the low-frequency component, giving the control wave included in Fig. 15.22. This type of oscillator is easily brought into step by the synchronizing impulse from the separator circuits.

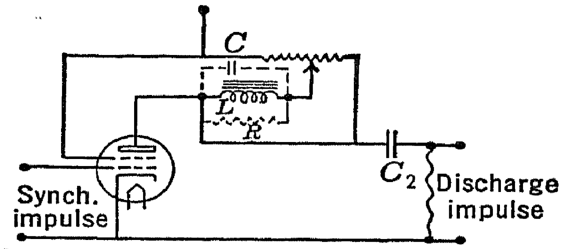


FIG. 15.21.—Dynatron Oscillator.

Various forms of multivibrators may be used to provide the required impulses. A conventional multivibrator circuit is illustrated in Fig. 15.35. In general this type of oscillator requires two vacuum tubes and therefore is not considered economical in a commercial receiver.

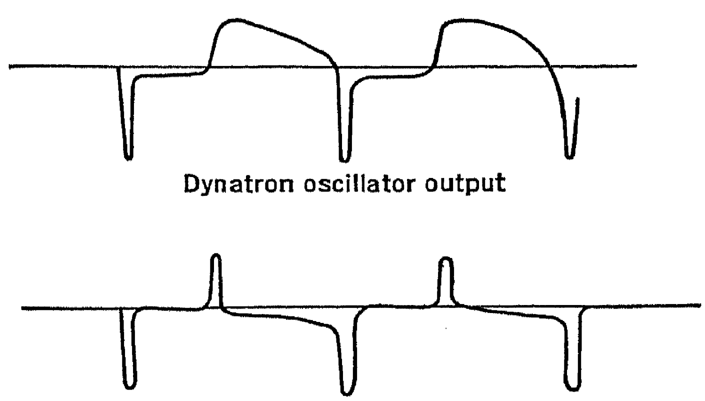


FIG. 15.22.—Wave Shape from Dynatron Oscillator.

Most modern receivers employ a blocking oscillator to obtain the necessary impulses. The arrangement of circuit elements for a typical blocking oscillator is shown in Fig. 15.23. If the resistor R has a low value, the circuit oscillates as a conventional feedback oscillator, the frequency being determined by the inductance and the distributed capacity of the transformer. The value of the circuit constants of the transformer are so chosen that this frequency is much higher than that of the impulse oscillations desired. With a large resistance across the condenser, as soon as oscillation starts, rectification by the grid builds up a negative voltage across the condenser to cutoff. Until the negative charge leaks

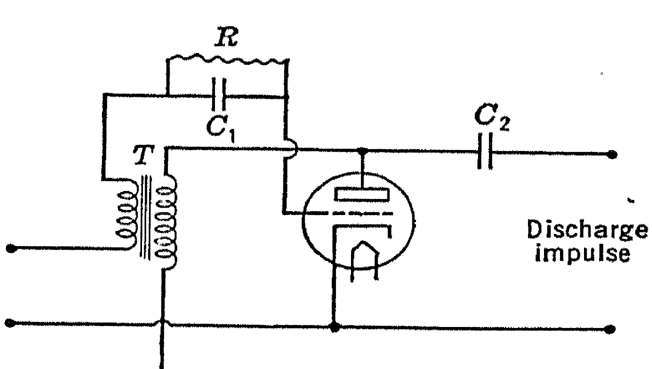


FIG. 15.23.—Blocking Oscillator.

off through R the circuit is paralyzed. As soon as the condenser is discharged, the cycle starts again. The wave forms which appear in various parts of the circuit are shown in Fig. 15.24 and serve to clarify the operation of this type of oscillator. The frequency of oscillation is chiefly dependent upon the time constant of the RC circuit. More exactly, however, the frequency depends in an extremely complicated way on all the circuit elements, including the tube. A quantitative analysis of the circuit is difficult because the differential equations describing its actions contain variable coefficients. A numerical solution can be obtained with the so-called method of isoclines.*

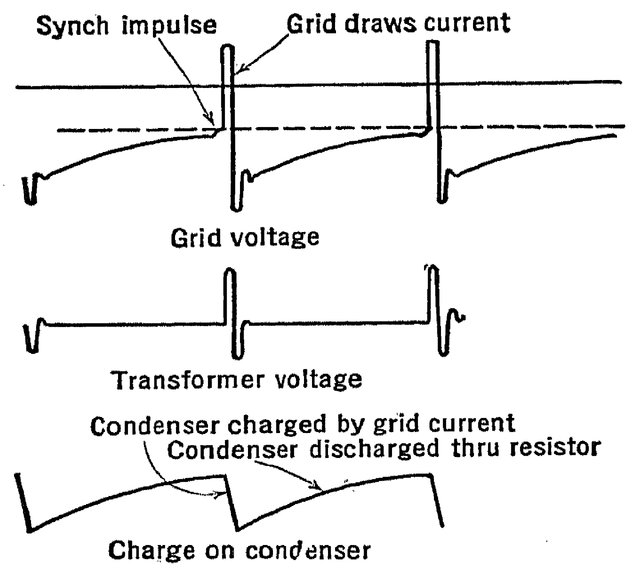


Fig. 15.24.—Wave Shapes Appearing in Blocking Oscillator.

This type of oscillator synchronizes easily over quite a range of frequency, without phase shift, when a small impulse is applied to the grid of the tube from the low end of the transformer. In most of its characteristics the blocking oscillator has been found to be very adequate as a discharge impulse generator.

15.6. Synchronization. The impulses which trigger the control oscillators of the horizontal and vertical deflection generator accompany the picture signal, being a part of the complete video signal. These synchronizing impulses must meet the following requirements:

1. They must not interfere with the picture.
2. They must be easily separable from the picture signal.
3. They must be such that vertical and horizontal control impulses can be distinguished electrically.
4. They must be capable of controlling the oscillator in the presence of ordinary interference.

* See F. Kirschstun, reference 12; and also I. G. Maloff, reference 13.

The first condition is met by placing the synchronizing impulses during the return time of the scanning beam, as was described in Chapter 7.

As was pointed out in the chapter mentioned, the synchronizing signals cannot be distinguished from the picture signal by their frequency content and therefore must be selected from the video signal by virtue of some other attribute. They are characterized by their time of occurrence and their amplitude. The time cannot readily be used as the means of selection, because it is the impulse which times the system itself. Therefore of necessity amplitude selection is used.

Separation of vertical and horizontal control impulses can be effected by means of differences of shape, duration, or amplitude of the two types of impulses. Choice of the most suitable means involves circuit considerations which will be discussed in a succeeding section.

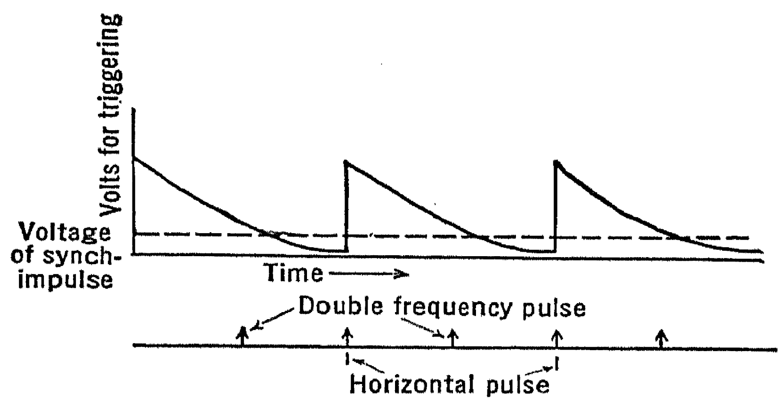


FIG. 15.25.—Sensitivity of Control Oscillator as a Function of Time.

The effect of interference upon synchronization depends upon the nature and amplitude of the interfering signals. If the amplitude of the synchronizing signal is sufficient to insure its positive separation from the picture signal, random noise will not interfere with synchronization unless its level is well above that which can be tolerated in the picture itself. However, sharp interfering pulses may cause the deflection generator to start prematurely. It is possible, of course, for the disturbance to oppose the synchronizing impulse and thus delay the start of the control oscillator, but this will be a rare occurrence. A disturbing pulse will only affect synchronization if its amplitude is comparable with the control signal itself, and then only if it occurs just before the end of a line or frame, at which time the control oscillator is easily started (see Fig. 15.25). The amplitude of the synchronizing impulses is made large enough so that interference which is not too obtrusive on the Kinescope screen will not destroy synchronization. On this basis the amplitude of the synchronizing impulse is made 20 per cent greater than the maximum picture signal. As far as immunity to interference and the ease of

separation are concerned, the polarity of the synchronizing impulse may be in the direction of either black or white. In other words, the signal may be "blacker than black" or "whiter than white." Since the former drives the grid of the viewing tube below cutoff during return time, some circuit simplification is effected by its use.

15.7. Separation of Video Signal and Synchronizing Impulse. The control impulses are selected from the complete video signal by a limiter

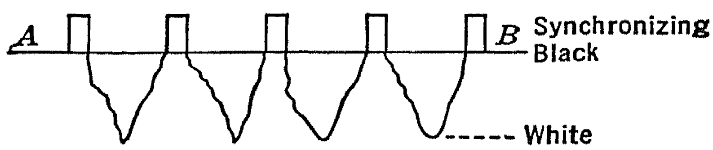


FIG. 15.26.—Synchronizing and Picture Components in Complete Video Signal.

tube. This is essentially a distorting amplifier which amplifies only that portion of the signal which in Fig. 15.26 is represented as lying above the line $A - B$.

A number of circuits are suitable for this purpose. One of the simplest is shown in Fig. 15.27. The tube is biased at such a point that the grid is just starting to draw current at the black level of the picture signal. A high-valued resistor is in series with the grid. When the signal is in the

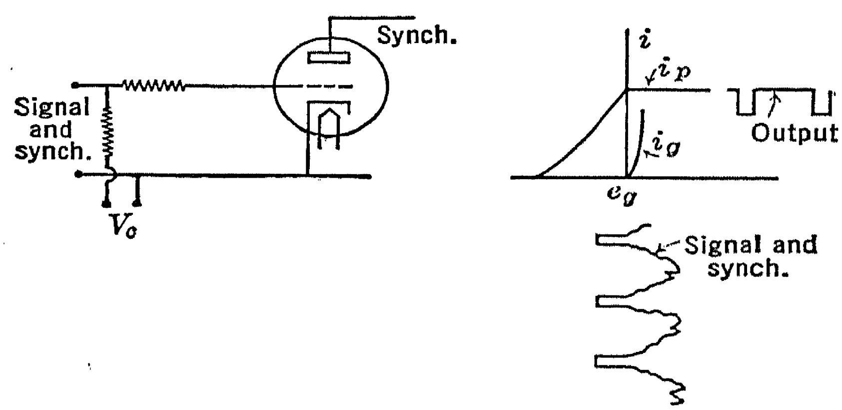


FIG. 15.27.—Positive-Grid Separator.

negative direction, as it is for the synchronizing impulse, the tube amplifies in the ordinary way; a signal of the opposite polarity, corresponding to the picture signal, causes the grid to draw current, producing a voltage drop through the resistor which opposes the signal. This type of limiter can be used only where the d-c level of the picture has been established. Furthermore, the shunt capacity of the tube causes a loss of high frequencies in the synchronizing impulse.

A more satisfactory limiter consists of a tube biased to cutoff, where the synchronizing impulse causes the grid to become positive, while the

picture signal swings it in the negative direction. If the video signal has sufficient amplitude this circuit can easily be made to select its own d-c level when supplied with an a-c picture signal. The circuit arrangement

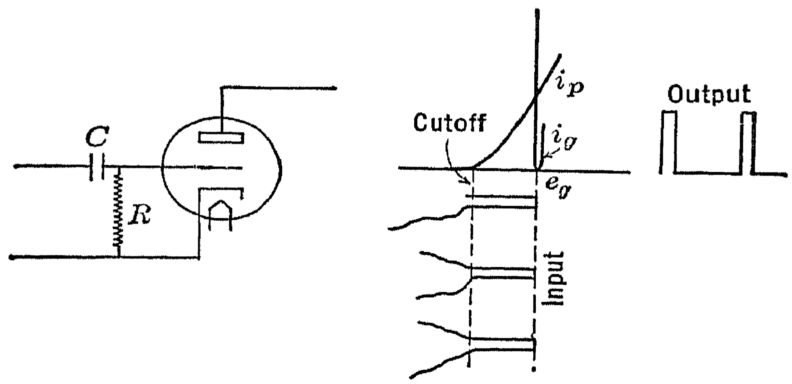


Fig. 15.28.—Grid-Cutoff Separator.

accomplishing this is shown in Fig. 15.28. The large grid resistor R is connected to the cathode so that, in the absence of signal, the grid is at zero bias. When a positive signal is applied to the grid through the condenser C, grid current flows into this condenser until the grid returns to cathode potential. Equilibrium is established very rapidly because even a small positive potential causes a large current to flow to the grid. This condition having once been reached, any signal more negative than this positive peak stops the flow of grid current and the condenser C can only lose its charge through the resistor. The time constant of the RC circuit is made large; consequently, the rate of discharge is low. Therefore the grid tends to maintain itself at such a d-c potential that zero bias corresponds to the maximum positive peaks of the signal, or in this case, the top of the synchronizing impulses. If the amplitude of the impulses is sufficient so that black corresponds to plate current cutoff, only the synchronizing impulses are amplified, the rest of the signal occurring beyond cutoff.

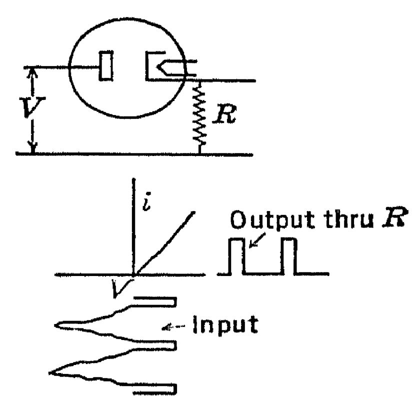


Fig. 15.29—Diode Separator.

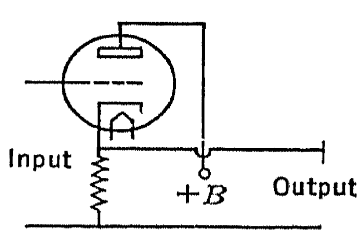


Fig. 15.30.—Cathode-Follower Separator.

Two other limiter circuits are shown in Figs. 15.29 and 15.30. The first is a biased diode rectifier; the second, a cathode follower limiter. Both these arrangements require a fixed d-c level.

15.8. Selection of Vertical and Horizontal Synchronizing Impulses. After the vertical and horizontal impulses have been isolated from the complete signal, there remains the problem of separating the two im-

pulses. This selection is effected by making the two impulses somewhat different in nature. A great many different types of impulses have been proposed, but in general they can be grouped into three classes.

The first includes those which can be separated by virtue of a difference in amplitude. A typical example is illustrated in Fig. 15.31. Separa-

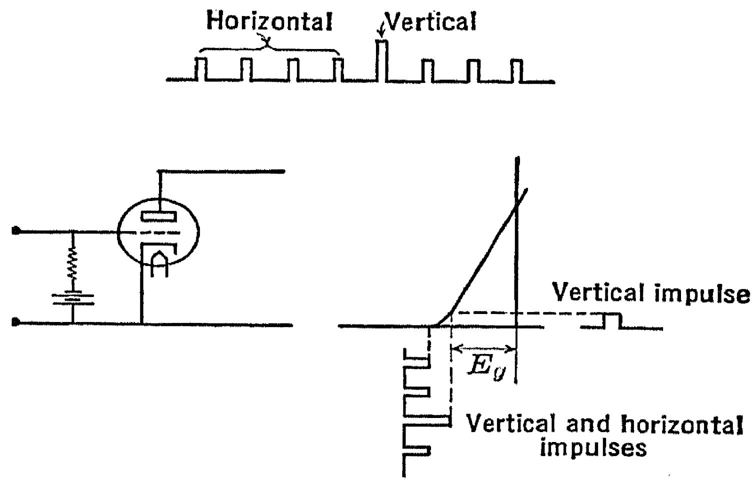


FIG. 15.31.—Amplitude Discrimination Selector.

tion is accomplished by means of limiter tubes in much the same way that synchronizing and picture signals are separated. The advantage of this type of signal is that it permits very accurate timing of the vertical oscillator. On the other hand, if the horizontal impulses are of sufficient amplitude to override interference, the large vertical pulse must be

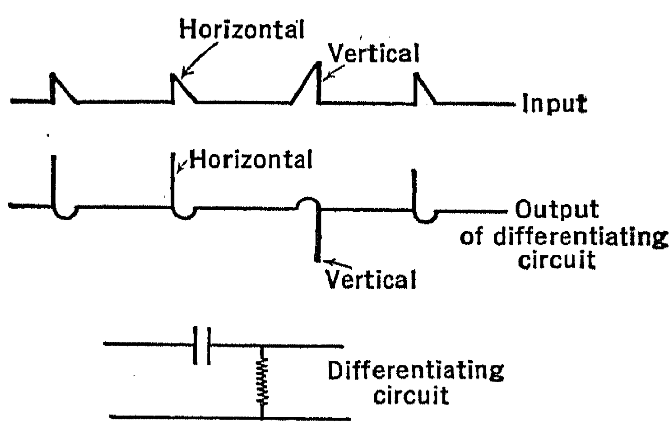


FIG. 15.32.—Differentiation Selector.

of much greater amplitude than is required to eliminate interference. Such vertical pulses therefore place a high peak power load on the transmitter.

Impulses of the second class are those which are distinguished by the slope or shape of their leading or trailing edges. A characteristic example is shown in Fig. 15.32. Selection in this case is made by differentiating the incoming signal by circuits of the type shown.

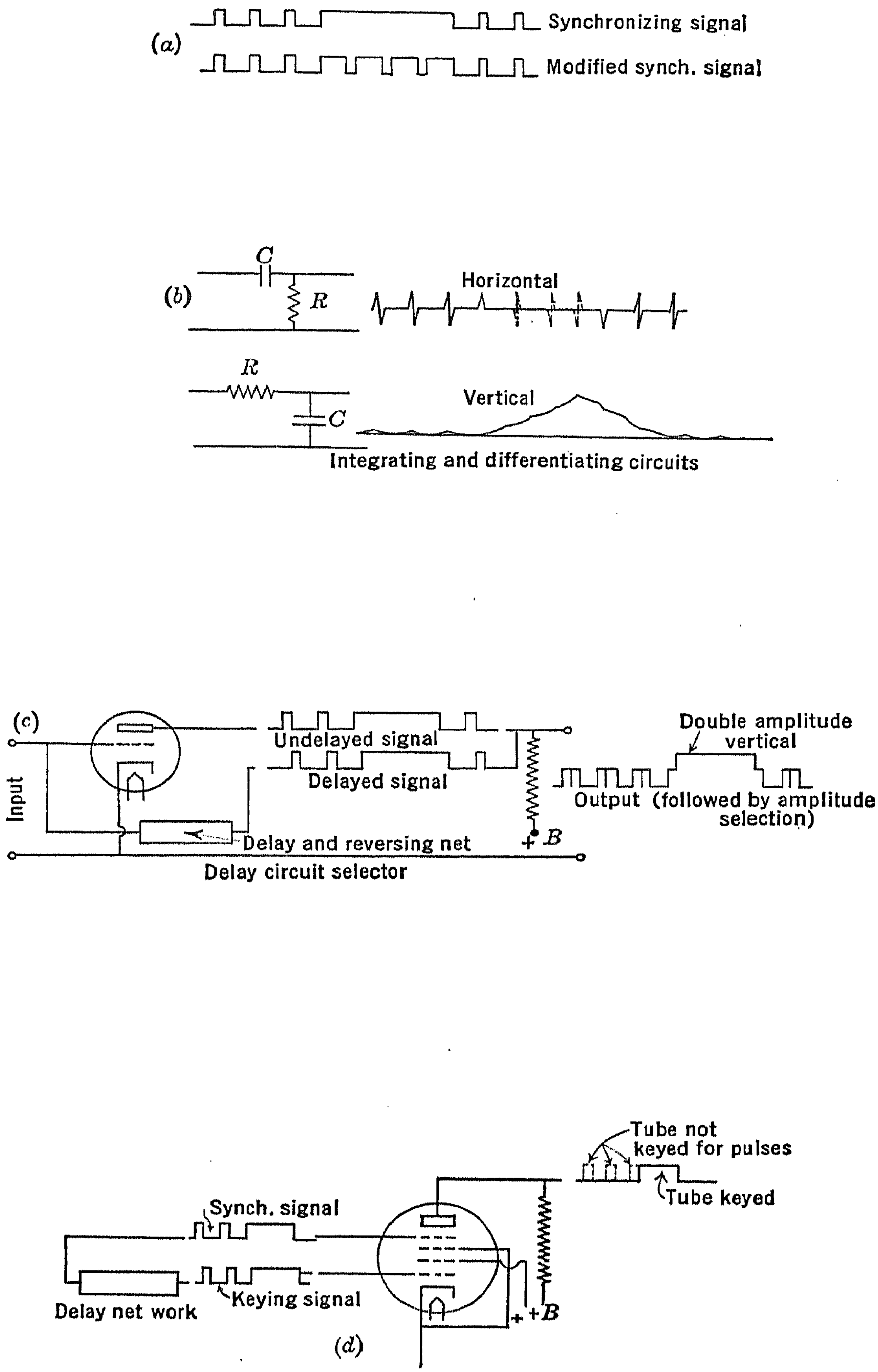


FIG. 15.33.—Duration Selectors.

The final class of synchronizing signals are those having vertical and horizontal impulses of different duration, the vertical being much longer than the horizontal, as illustrated in Fig. 15.33*a*. Selection can be made in a number of ways. For example, a differentiating circuit will respond to the short impulses, whereas an integrating circuit will respond to the long impulses. Such circuits are shown in Fig. 15.33*b*. The circuits shown in Figs. 15.33*c* and *d* use delay networks.

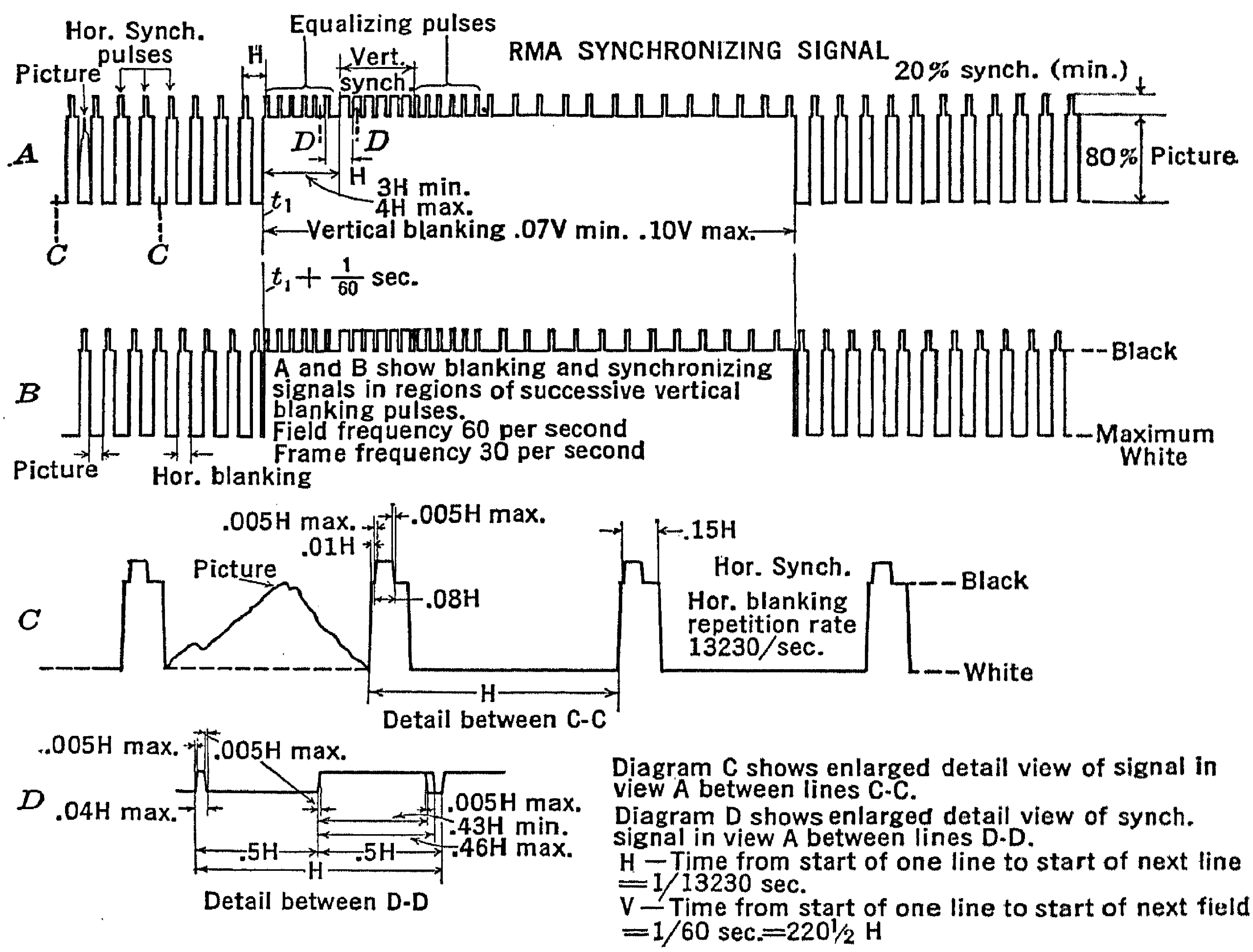


FIG. 15.34.—RMA Standard Television Synchronizing Signal.

Signals of the third class have been adopted as standard for this country by the Radio Manufacturers Association. The actual synchronizing signal was illustrated in Chapter 7, and is reproduced again at this point for convenience. The horizontal signals consist of nearly rectangular impulses of short duration. The vertical impulse is rectangular and has a duration of three line periods. In order to maintain horizontal synchronization the vertical impulses are serrated. Actually the serration occurs at twice line frequency, but since the horizontal oscillator is insensitive at the half periods it responds only at line frequency. For about three

lines before and three after the vertical signal, the horizontal impulses occur at double frequency. This also has no effect on horizontal synchronization, but makes the vertical signals for the odd and even field traversal of the interlaced pattern identical as far as an integrating circuit is concerned. Such similarity is quite essential since the vertical synchronization must be accurate in order to maintain interlacing. If ordinary horizontal pulses continued up to the vertical impulse, the difference between the odd and even field signals would be enough to spoil the timing of the oscillator. For certain commonly used circuits (e.g., the blocking oscillator and hard discharge tube described), the integrated area of the vertical impulse influences the change in voltage upon discharge, and consequently the size of the vertical deflection. Therefore the double frequency impulse is continued after the vertical impulse, making the integrated vertical pulse for odd and even field traversals alike.

15.9. Formation of the Complete Synchronizing Signal. The type of synchronizing signal having been selected, there still remains the problem of forming the impulses and adding them to the picture signal.

The impulses are formed by a chain of suitable multivibrator oscillators and shaping circuits. These impulses are added together and to the picture signal by means of keying or mixing tubes. There are many chains which are satisfactory, and the choice depends upon the particular installation. Only the broader aspects can be taken up here.

The highest impulse frequency is twice line frequency. Therefore there is a primary oscillator delivering pulses at a rate of 26,460 per second. From this a multivibrator, adjusted to respond to every other pulse, produces 13,230 pulses per second. Another chain of multivibrators responds to 7, 7, and 9 impulses (i.e., $7 \times 7 \times 9 = 441$), giving a 60-cycle impulse for the vertical signal. As has already been mentioned, the synchronizing chain is controlled by the frequency of the commercial power serving the transmission area. This control can be accomplished by adding the 60-cycle impulse to the a-c power, rectifying the sum with a peak voltage rectifier, and using the output to govern the frequency of the primary oscillator.

The multivibrator for the required frequency division is a form of relaxation oscillator. A typical circuit is shown in Fig. 15.35. This type of oscillator, like the previously described blocking oscillator, is sensitive to a control impulse just before its period of activity, and therefore can easily be made to lock in on some multiple of a control frequency. Its output is essentially in the form of impulses whose duration is governed by the circuit constants. A limiter tube can be used where further shaping of the signal is necessary.

The keying tube used in the addition of the impulses is merely a double-control grid tube. The impulses to be added are applied to one grid, and the other grid is maintained at cutoff until a keying signal makes this grid positive and allows the impulses to be amplified.

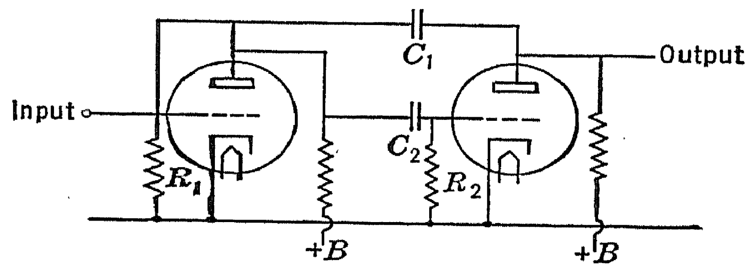


Fig. 15.35.—Multivibrator.

The building of the synchronizing signal using these elements is diagrammed in Fig. 15.36. The net consists of suitable multivibrators, keying tubes controlling the passage of impulses, and forming circuits. The forming circuits are in reality special multivibrators, actuated by the 60-cycle oscillator, and so adjusted as to produce 60-cycle impulses of the correct duration and timing to serve for the required keying.

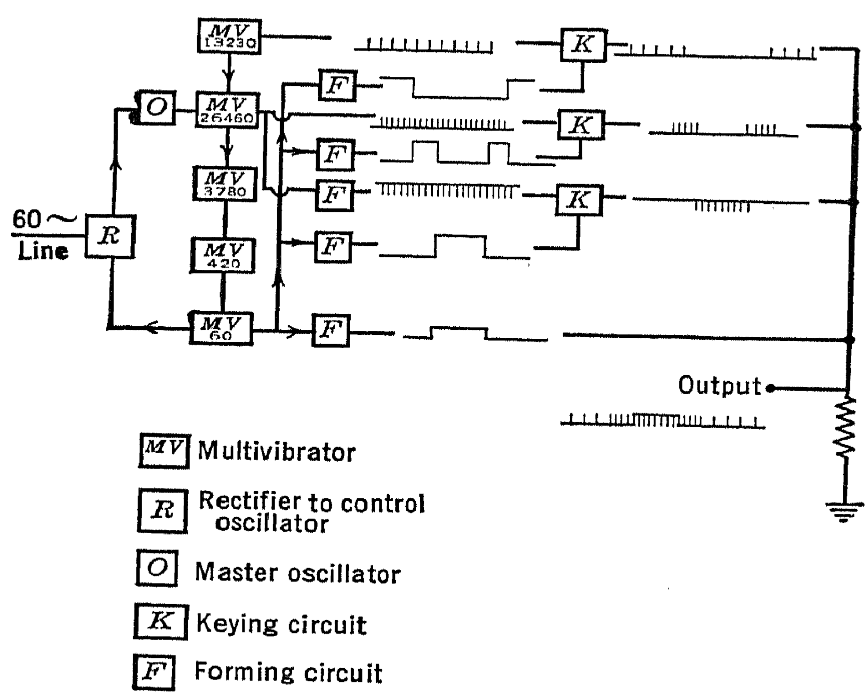


Fig. 15.36.—Generation of Synchronizing Signal.

The actual circuits in a practical television installation are more involved, including many refinements for increasing the stability of operation. For example, considerable advantage is to be had if one oscillator provides the leading edge of all impulses, long or short.* Such a synchro-

* Patent applied for, A. V. Bedford.

nizing signal generator avoids the almost inevitable shifts in phase which are present when changing from one type of impulse to another.

15.10. Special Problems of Scanning. *Interlaced Scanning.* Two methods of interlacing are commonly employed, one for which the complete pattern consists of an even number of lines, the other where the pattern has an odd number of lines.

The even-line interlacing is obtained by shifting the level of the alternate sawtooth waves producing vertical deflection by a distance equal to one horizontal line. Thus the alternate field patterns are displaced with respect to one another by the width of one line. The vertical sawtooth wave form is shown in Fig. 15.37.

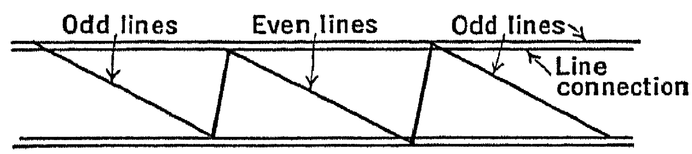


FIG. 15.37.—Sawtooth Wave for Even Interlacing.

Odd-line interlacing, the type adopted by the Radio Manufacturers Association, uses vertical waves which are all identical. However, the first vertical sweep ends when a horizontal line is only half way across, and the second sweep, starting at exactly the same height as the first, begins with the horizontal deflection at the mid-point. Thus, line number $N/2 + 1/2$ is shifted by the distance of one line from line number 1 (i.e., shifted half the distance of the double-spaced lines laid down by the first sweep).

The first method of obtaining interlaced scanning requires very accurate shifting of the vertical wave, which is difficult to attain with simple circuits. In addition, the vertical deflection amplitude must be extremely constant. The second method also requires very accurate deflection in the vertical direction, but has the advantage of identical vertical waves for every field sweep.

Keystoning. The mosaic of the Iconoscope makes an angle of 60° to the axis of the gun. Therefore if the scanning pattern is formed by vertical and horizontal deflections having constant angular amplitude, the pattern will not be rectangular. Since the bottom of the mosaic is closer to the gun than the top, the top lines will be longer than the bottom lines, and the area swept out on the screen will be trapezoidal, or keystone, in shape. This will be clear from Fig. 15.38*a*.

To correct for this it is necessary to modulate the horizontal deflection with vertical sawtooth wave. In this way a decrease in angular amplitude at the top and an increase toward the bottom can be made to overcome

the trapezoidal distortion in the pattern on the slanting screen. The appearance of the modulated sawtooth wave is shown in Fig. 15.38*b*.

This modulation can be carried out in a number of ways. Essentially the gain of the deflection amplifier must be linearly increased periodically at vertical frequency, and the vertical sawtooth wave component, present in addition to the modulated horizontal sawtooth, must be removed. An alternative is to cause a periodic increase in amplitude of the

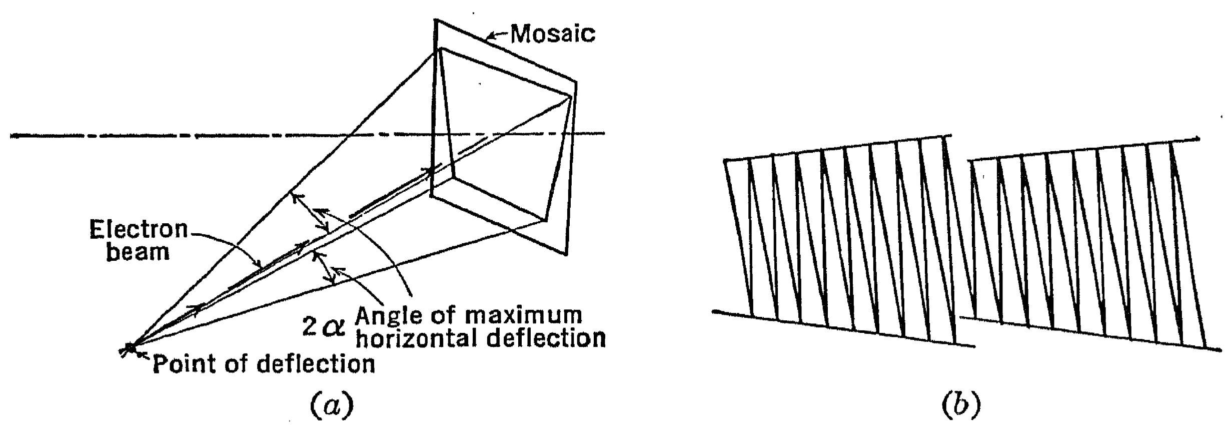


FIG. 15.38.—Origin and Correction of Keystone Distortion.

sawtooth generator itself. Actually the first method has been almost universally adopted, and it will be the only one discussed. Two systems of modulation only are treated, but it will be obvious that there are a great many other possibilities.

The first modulator employs a variable-mu horizontal deflection amplifier tube. The grid bias of this tube is varied periodically by the vertical sawtooth. This produces a modulated horizontal sawtooth, plus a vertical component due to the variation in bias. The second component is removed by applying an inverse vertical sawtooth of amplitude equal to that present in the output of the modulator. The circuit is shown in Fig. 15.39.

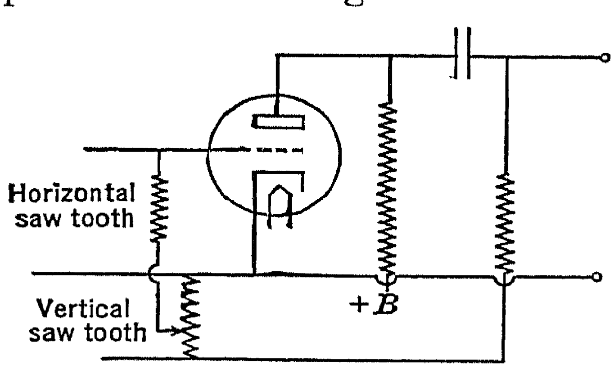


FIG. 15.39.—Simple Keystone Correcting Circuit.

The second system is the so-called balanced modulator system. Here two modulator tubes are used. These tubes are of the double-control grid type (hexodes), the horizontal frequency being put on one grid of each tube in inverse phase, while the modulating vertical frequency is applied to the other grid in like phase. The outputs are coupled in push-pull to a transformer. Fig. 15.40 will make this arrangement apparent.

Both these systems have their particular applications, and each has its advantages and disadvantages.

The keystone effect also produces a non-uniformity in the distribution of horizontal lines. To avoid this non-uniformity, the rate of increase in the vertical deflecting wave must be less at the top than at the bottom. If this correction is made it is also necessary to change the modulation of the horizontal deflection in like manner. These corrections can be made by appropriately choosing the value of the resistor in series with the charging condenser of the vertical sawtooth generator.

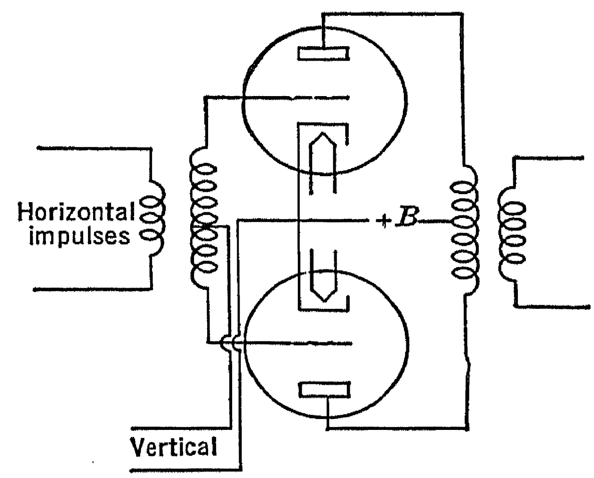


Fig. 15.40.—Balanced Modulator for Keystone Correction.

Interference and Scanning. The three main types of interference to be guarded against when designing a scanning system are: (1) effect of noise and static upon the synchronizing mechanism; (2) crosstalk between vertical and horizontal deflection; and (3) alternating-current ripple or hum from the power supply feeding through to the deflection. There are, of course, many other less important causes of interference.

The effect of interference on synchronizing has already been pointed out. Unless the disturbances are of such a nature and magnitude as to spoil the picture itself, the only effect is the occasional loss of a line or of interlacing.

Crosstalk is the term applied when the deflection in one direction interacts with that at right angles to it. It is the result of inadequate electrical isolation between the two deflecting systems, or a poorly designed deflecting yoke. The first may be the result of insufficient shielding of the deflection amplifiers, or leads, or of coupling through a common power supply. The second, involving the design of deflecting coil and yoke, is beyond the scope of this chapter.

Where the vertical deflection intermingles with the horizontal, the result is a distortion of the pattern shape. If the complete sawtooth is present as an addition, the pattern will be sheared; if it is a modulation, the shape will be keystoned. Usually only some of the harmonics interfere, which results in crooked sides to the scanned area.

Harmonics of the horizontal sawtooth appearing in the vertical deflection cause the scanned lines to be wavy instead of straight. Since the Iconoscope is sensitive to the type of scanning pattern, this second type

of distortion is very serious at the transmitter as it may result in a strong spurious signal. At the receiver its principal effect is to spoil the perfection of the shape of the image and interfere with interlacing.

One of the reasons for using 30-frame picture transmission is that it can be synchronized with the power supply, and thus minimize the effect of hum, not only on the picture but also on the scanning pattern.

Hum may either modulate or add to the deflection in either direction. Modulation of the horizontal deflection by 60-cycle interference will cause sides of the pattern to have a sine-wave shape, rather than straight, the opposite sides being mirror images of one another, while a hum addition will cause the sides to have a similar wave appearance, but the sides will be parallel. The effect of 60-cycle hum on an interlaced pattern with

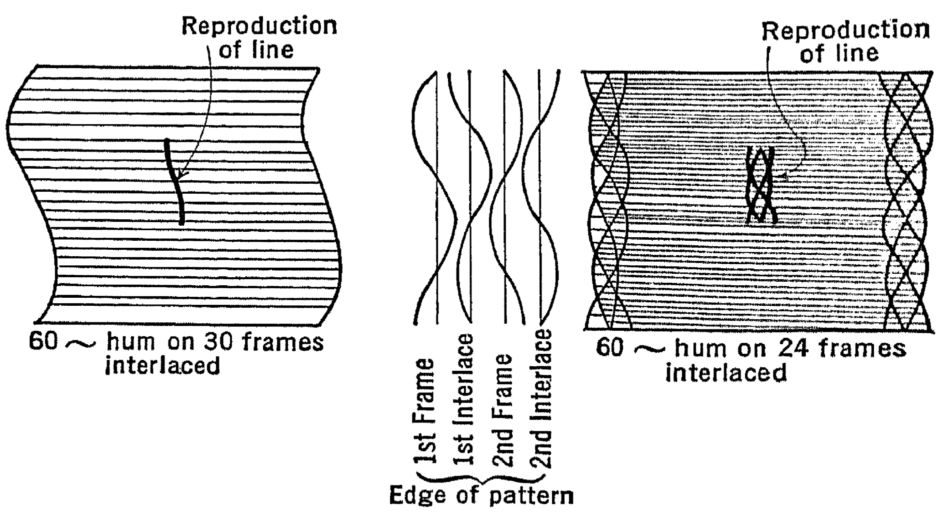


Fig. 15.41.—Effect of Hum on Scanning Patterns.

24 frames per second is much more serious than on one with 30 frames per second, as can be seen from the comparative patterns given in Fig. 15.41. It will be immediately evident that a small amount of hum in the 30-frame picture will not be very objectionable, whereas in the 24-frame one it will result in a blurring or even a doubling of the picture. Where the scanning rate and hum differ by only a few cycles per second the picture appears to sway—an effect which is very objectionable.

The effect of hum on the vertical deflection is to produce a sinusoidal variation in the density of distribution of lines. With a 30-frame picture, 60-cycle hum does not disturb the interlacing at all, but for a 24-frame picture the lines draw together in pairs and may even coincide, reducing the resolution of the picture.

If perfect filtering of the power supplies could be simply effected, the problem of 60-cycle interference would not, of course, have arisen, and in all probability a 24-frame picture would have been adopted to conform with moving-picture practice. However, in any practical receiver,

where cost of production is a matter of importance, there will be a small residual of hum, entirely unnoticeable where 30 frames are used, but materially affecting the resolution and quality of a 24-frame picture.

REFERENCES

1. I. G. Maloff and D. W. Epstein, "Electron Optics in Television," McGraw-Hill, New York, 1938.
2. F. Schröter (editor), "Fernsehen," J. Springer, Berlin, 1937.
3. J. C. Wilson, "Television Engineering," Pitman, London, 1937.
4. V. K. Zworykin, "Experimental Television System and Kinescopes," *Proc. I. R. E.*, Vol. 21, pp. 1655–1673, December, 1933.
5. R. D. Kell, "Description of Experimental Television Transmitting Apparatus," *Proc. I. R. E.*, Vol. 21, pp. 1674–1691, December, 1933.
6. G. L. Beers, "Description of Experimental Television Receiver," *Proc. I. R. E.*, Vol. 21, pp. 1692–1706, December, 1933.
7. R. D. Kell, A. V. Bedford, and M. A. Trainer, "Experimental Television System—Part II, Transmitter," *Proc. I. R. E.*, Vol. 22, pp. 1246–1265, November, 1934.
8. R. S. Holmes, W. L. Carlson, and W. H. Tolson, "An Experimental Television System—Part III, The Receiver," *Proc. I. R. E.*, Vol. 22, pp. 1266–1285, November, 1934.
9. P. T. Farnsworth, "Television by Electron Image Scanning," *J. Franklin Inst.*, Vol. 218, pp. 411–444, October, 1934.
10. I. G. Maloff, "The Cathode Ray Tube in Television," *Proc. Radio Club Amer.*, Vol. 12, pp. 31–36, October, 1935.
11. O. S. Puckle, "A Time Base Employing Hard Valves," *J. Television Soc.*, Series II, Vol. 2, Part V, pp. 147–155, June, 1936.
12. F. Kirschstun, "New Methods of Graphical Treatment of Relaxation Oscillations," *Arch. Elektrotech.*, Vol. 24, p. 731, 1930.
13. I. G. Maloff, "Solution of Vacuum Tube Problems by the Isocline Method," *Broadcast News*, No. 10, pp. 14–17 and 50, February, 1934.
14. F. Banneitz, "Standardization of Synchronizing Signals of the German Television System," *Telegraphen-Fernsprech und Funk Technik*, Vol. 27, p. 157, May, 1938.
15. D. v. Oettingen, R. Urtel, and G. Weiss, "On Single Channel Synchronizing in Television," *Telegraphen-Fernsprech und Funk Technik*, Vol. 27, pp. 158–166, May, 1938.
16. A. D. Blumlein, "The Marconi-EMI Television System—Part I, Transmitted Waveform," *J. I. E. E.*, Vol. 83, pp. 758–766, December, 1938.
17. R. D. Kell, A. V. Bedford, and M. A. Trainer, "Scanning Sequence and Repetition Rate of Television Images," *Proc. I. R. E.*, Vol. 24, pp. 559–576, April, 1936.

CHAPTER 16

THE TELEVISION TRANSMITTER

The difference between the type of transmitter required to radiate high-definition television signals and that used for sound broadcasting is so great as to be almost fundamental. The reason for this difference is the tremendous frequency band needed for picture transmission. Not only does the wide frequency band present a technical obstacle in itself, but it also makes necessary the use of ultra-high frequencies for picture transmission, as was pointed out in Chapter 7. The region of the frequency spectrum lying between 40 and 100 megacycles has been selected as suitable for television broadcasting, and seven 6-megacycle bands have been assigned therein for this purpose by the Federal Communications Commission. In Europe and elsewhere, similar bands have been given over to video transmission. The subject matter of the present chapter is concerned with television broadcast transmitters operating in this frequency range.

Like the transmitter used for sound broadcasting, the ultra-high-frequency transmitter consists of a carrier generator, a modulator, power amplifiers, a transmission line carrying the modulated carrier to the antenna, and finally the antenna itself. In principle these elements are similar to those used in broadcast transmission, but because of the extreme frequencies involved, problems which in broadcast transmission are only of secondary importance assume the proportions of major considerations. At the frequencies used for ultra-short-wave transmission, the series impedance due to the inductance of every lead and the shunt capacity between every electrode, or lead, and ground, must be taken into account. In fact, every lead must be considered as a transmission line whenever its length is comparable with that of the wave traveling over it.

16.1. Character of the Transmitted Signal. The modulation of an ultra-high-frequency carrier by the video signal to form the radio-frequency signal was introduced in Chapter 7. The modulating video signal may be of either polarity, giving rise to either positive or negative modulation. With positive modulation the amplitude of the radio frequency is a maximum for white portions of the picture while with nega-

tive modulation the converse is true, that is, the black regions of the picture increase the radio-frequency amplitude.

As has already been pointed out, practical considerations dictate the use of synchronizing impulses in the direction of a black signal. Therefore if positive modulation is used the synchronizing signal causes the radio-frequency signal to drop to a very low value (usually zero amplitude), while negative transmission results in a maximum signal for the synchronizing impulses.

The choice of positive or negative transmission is not very clear-cut, as there are reasons favoring each. If the effect of interference is neglected they are almost equally suitable, except that with negative transmission a reference for automatic gain control is more easily obtained. Interference, however, affects the two systems somewhat differently. The most common type of interference consists of pulses of short duration and high amplitude. The complete destruction of the synchronizing by severe interference is equally probable for either positive or negative transmission. However, if the interference is sufficiently severe to cause complete loss of synchronization, it will, in general, render the picture valueless even if synchronization is artificially maintained. Such interference, when not sufficiently severe to totally destroy the picture, occasionally can cause the impulse generator of the deflecting circuits to operate too soon. This occurs when a disturbing pulse arrives just ahead of the normal synchronizing impulse, at which time the deflection generator is extremely sensitive to any incoming signal. Negative picture transmission is somewhat more sensitive to this type of interference than positive modulation. This is because to interrupt synchronization in positive transmission the interfering signal must be of such amplitude and phase that it just cancels the incoming signal. If the phase of the interference is not exactly opposite to that of the carrier it will not cause the signal to become zero, and therefore will not interrupt the deflection. Likewise, if the interference is in phase opposition but its amplitude is greater than that of the video signal, it will result in over-modulation rather than loss of signal. Negative transmission, on the other hand, is susceptible to any impulse occurring near the end of a line or frame which causes the signal to rise to a value equal to, or greater than, that required for synchronization. This advantage of positive modulation is offset by the fact that interference which causes an increase in signal (either by direct addition or overmodulation) produces a white flash on the viewing tube screen. A corresponding interfering signal in negative transmission produces a dark spot. Bright flashes on the screen are much more objectionable than dark spots of similar size.

Negative picture transmission is at present favored in this country;

positive transmission is widely used abroad. The final choice can be made only as the result of extensive tests under actual working conditions.

The signal, whether positively or negatively modulated, can be radiated from the transmitter in the form of either a-c or d-c transmission. A-c transmission, as the name implies, means that the video signal supplied to the modulator has been amplified by an amplifier passing only its alternating-current components. Because of this, the signal at the modulator represents the voltage variation about an axis or reference voltage which is the time average of the signal. Thus, if the picture consists of a narrow white line on a black background, the carrier envelope for negative modulation will vary from approximately half value for the background to zero carrier at the white band, while a black line on a white background leads to a carrier which has approximately half value for the white background to maximum value at the black line. It will be apparent that the instantaneous magnitude of the r-f signal has therefore no real meaning in terms of picture brightness. However, the complete signal carries the information which determines the correct light level of the picture, for the pedestal which accompanies the synchronizing signal always corresponds to black. Relatively simple circuits at the receiver can, by making use of the pedestal, establish the correct viewing-tube bias, as will be explained in the next chapter.

The video signal, on the other hand, may be amplified up to modulation level by means of a d-c amplifier or its equivalent. If this is done the magnitude of the r-f signal has absolute significance in terms of picture brightness. For negative modulation a black area, irrespective of size, results in maximum carrier amplitude (exclusive of the synchronizing signal), while a white area leads to zero or minimum carrier. The carrier envelopes for a bright line on a black background and a dark line on a bright background are shown for a-c and d-c transmission in diagrams (*a*) and (*b*), respectively, of Fig. 16.1.

Considerable economy in the use of any given transmitting equipment can be effected with d-c transmission. This saving will be made clear by reference to Figs. 16.1*c* and *d*. A given transmitter has available a certain maximum voltage range for the carrier envelope. This range, for convenience, is referred to the voltage applied to the modulator and is shown in Fig. 16.1*c* by the double-headed arrow whose ends represent maximum and minimum carrier. The reference axis about which the video signal voltage swings in a-c transmission must be at or near the midpoint of the modulator voltage range in order to transmit the two types of picture signal shown in the diagram. It will be seen that the voltage change from black to white is only about one-half the total avail-

able voltage swing. On the other hand, if a d-c video signal is applied to the same modulator, a "black" signal will correspond to a maximum positive voltage, while white corresponds to a maximum negative voltage; hence the full voltage swing of the modulator is available for the signal voltage. In other words, the useful voltage range with a-c transmission is only half as great as it is with d-c transmission. Actually the power obtainable by d-c transmission is not four times as great as by a-c transmission, which might be expected from the above discussion. In the first place the synchronizing signal and pedestal which occupy about 20 per

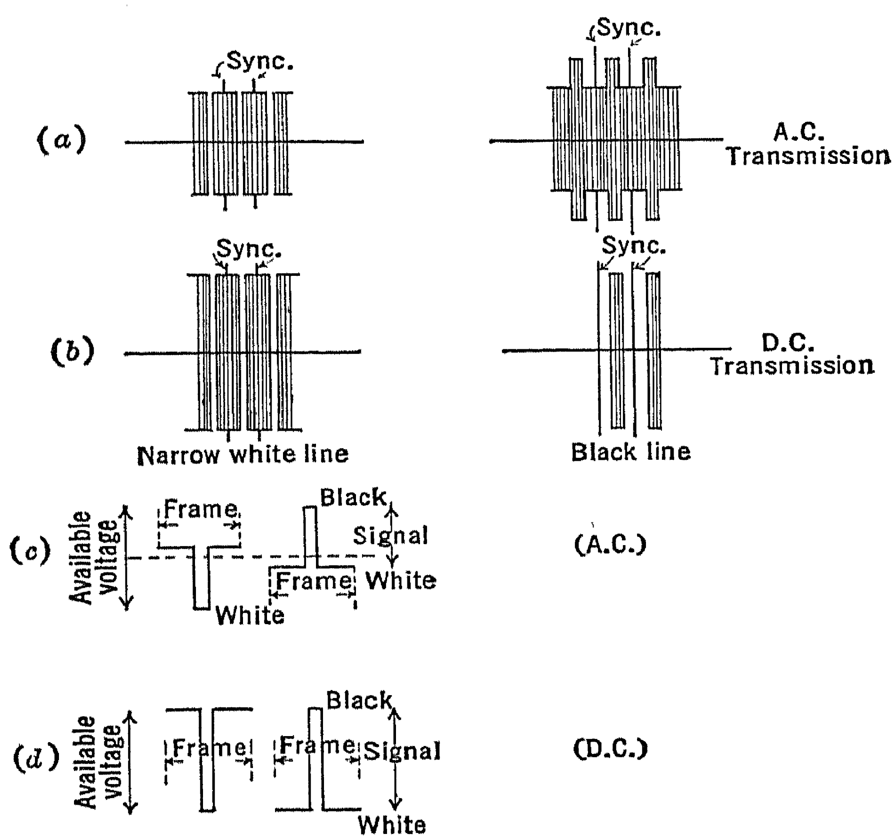


Fig. 16.1.—A-C and D-C Transmission of Video Signals.

cent of the transmission time reduce the power saving. (*Note:* For d-c transmission, synchronizing must always produce maximum carrier, while for certain types of pictures with a-c transmission it may be only slightly more than half maximum carrier.) Furthermore, the transmitter is limited not only by maximum voltage range but also by average power dissipated. The actual gain in power by the use of d-c transmission is between 1.5 and 2.5 times.

A carrier modulated in the ordinary way can be resolved into the carrier and two symmetrically placed sidebands. The spectrum will then appear as shown in Fig. 16.2*a*, which corresponds to a carrier modulated with a video signal containing all frequencies up to 3 megacycles. The position of the audio band relative to the video signal is also indicated

in this figure. This band is not related to the video carrier, but has its own carrier, placed at a fixed distance from the video carrier.

In Chapter 7 it was shown that a single sideband, together with the carrier, contains *all* the information necessary to synthesize the picture, and that considerable advantage is obtained if only a single sideband is used. This being so, obviously it would be highly desirable if only a carrier and a single sideband, as illustrated in Fig. 16.2*b*, were generated by the transmitter. Not only would this lead to a doubling of the effective power of the transmitter, but also it would reduce the space in the frequency band required to transmit a given picture.

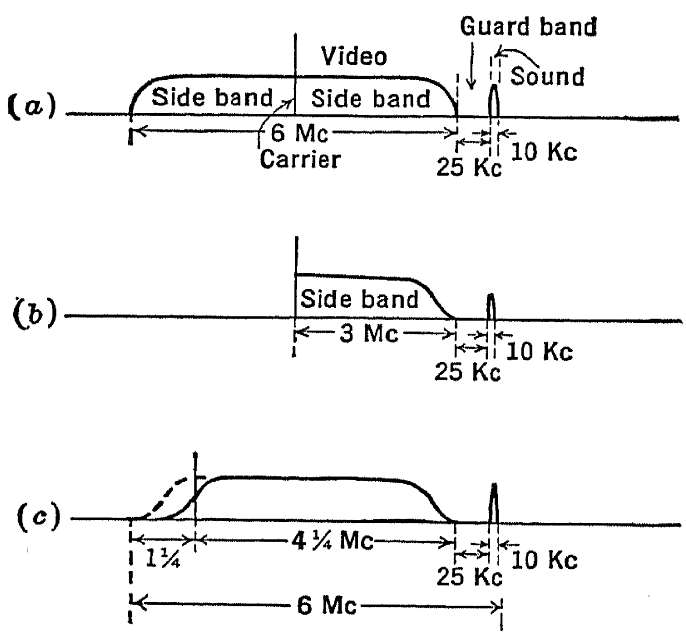


FIG. 16.2.—Frequency Distribution for Transmitters Radiating Single and Double Sideband Signals.

Unfortunately at the present state of the art it is not possible to make an efficient modulator and power amplifier which will supply only the wanted carrier and single sideband to the antenna. As will become more apparent as the discussion proceeds, this is due to the fact that the tuned circuits which must be used have phase characteristics which invariably lead to inefficient amplification unless the carrier coincides with their resonant frequency.

A partial improvement can be effected, however, by filtering out one sideband from the signal by means of filters between the transmitter and the antenna. The transmitter in this case generates both sidebands, and the power contained in one sideband is dissipated in the filter. Therefore no saving in power is to be had. The economy in utilization of the available frequency band is retained.

Even this compromise cannot be carried out in full because of filter

limitations. If a filter could be designed which had a suitable cutoff almost at the carrier, and no phase distortion over the pass band, so that the resulting signal was of the form shown in Fig. 16.2*b*, faithful picture reproduction could be obtained. However, practical filter considerations have made it advisable to transmit between 1 and 1½ megacycles of the unwanted sideband. The final shaping of the pass band takes place in the receiver. Though the saving is not as great as if this expedient were not necessary, nevertheless there is a gain of nearly 50 per cent in frequency utilization.

To summarize: the signal which is considered most favorable at present is one which is negatively modulated, with "blacker than black" synchronization, and contains the d-c component. One sideband should be partially suppressed so as to permit a wider video band to be transmitted within a given ultra-high-frequency channel. Finally, the carrier frequency lies between 40 and 100 megacycles.

16.2. The Carrier Generator. The carrier generator supplies a constant-amplitude radio frequency to the modulator. The power level at which this carrier is supplied depends on the nature of the modulator and the power amplifier stages which follow it. Usually the carrier level at the modulator is quite high in order to avoid the necessity of subsequent amplification of the modulated carrier.

The principal elements making up the carrier generator are a master oscillator, frequency multipliers where required, and a power amplifier.

The two types of oscillators most frequently employed in the carrier generator for ultra-high-frequency work are the piezoelectric crystal oscillator, and the "high-Q" tank circuit oscillator.

Crystal oscillators are common in broadcast practice. In essence a crystal oscillator is a sharply resonant mechanical vibrator with coupling between an electric circuit and the mechanical motion. The piezoelectric effect present in certain crystals provides this coupling. Quartz crystals have proved most satisfactory for this purpose.

A natural quartz crystal exhibits threefold symmetry about an axis known as the optic axis. This optic axis is electrically inactive. However, along the axes lying in a plane normal to the optic axis, quartz manifests a marked piezoelectric effect. Fig. 16.3 shows a crystal projected in a plane perpendicular to the optic or z-axis. In this figure the three axes Y_1, Y_2, Y_3, normal to the three main faces, are known as the mechanical axes, while the axes X_1, X_2, X_3, at right angles to the mechanical axes, are the electric axes. If the crystal is strained along one of the three mechanical axes, charges will be induced on the surfaces of the crystal normal to the electric axis which is at right angles to the direction of strain. Similarly, an electric field applied parallel to an electric axis will

produce a deformation along the mechanical axis perpendicular to the field.

If a rectangular segment is cut out of a quartz crystal in such a way that its face includes the Y and Z axes, as shown in Fig. 16.4a, and an alternating field is applied normal to the face, mechanical vibration will occur in the Y direction. As the frequency is varied a rate of vibration can be found which corresponds to the mechanical resonance frequency of the Y direction of the rectangle. At this frequency a field of small amplitude can sustain a very considerable mechanical vibration. Furthermore, if the electrical oscillations are stopped, the crystal will continue to vibrate at this frequency for a large number of cycles owing to its elastic properties, inducing alternating charges on its two faces. In other words, the electrical behavior of the crystal plate

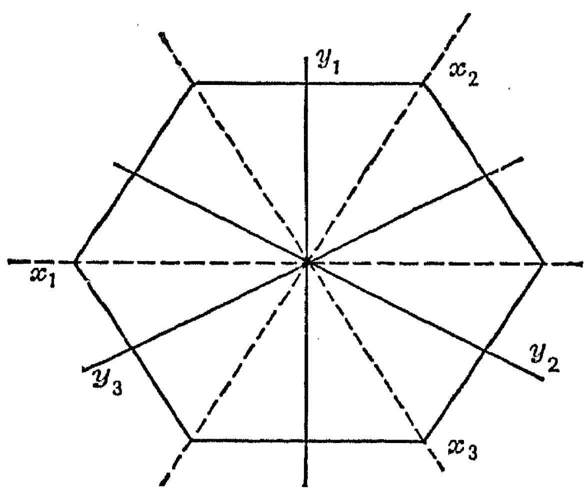


FIG. 16.3.—Orientation of Electrical and Mechanical Axes in Quartz Crystal.

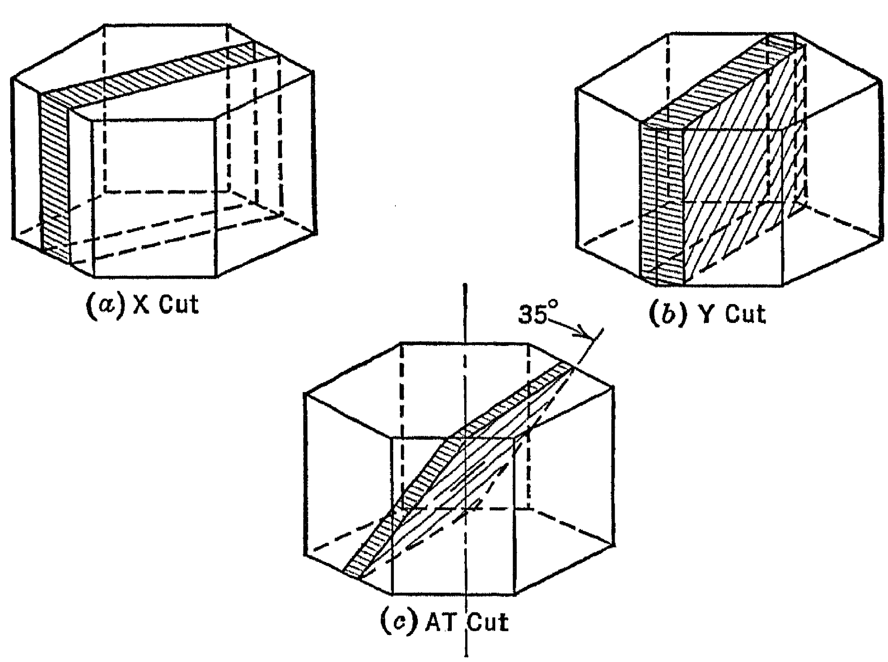


FIG. 16.4.—Cuts Yielding Oscillator Crystals.

is exactly the same as that of a very slightly damped (high-Q) parallel resonant circuit.

Under the resonant conditions just described, longitudinal mechanical standing waves whose motion is in the Y direction are set up. Similar

standing waves can be established normal to the YZ face which will correspond to a much higher frequency.

Useful crystals can also be cut to include other crystallographic directions. A slab cut so that its face includes the X axis and makes an angle of about 35° to the optic axis results in a crystal whose resonant frequency is practically independent of temperature. In addition, a crystal so cut will be almost entirely free from parasitic oscillations which result from coupling between the various crystallographic directions. Such a cut is known as an AT cut. Other orientations of the crystal plate are also employed in practice to obtain high stability. The V-cut crystal, for example, is another temperature-independent crystal which finds frequent application.

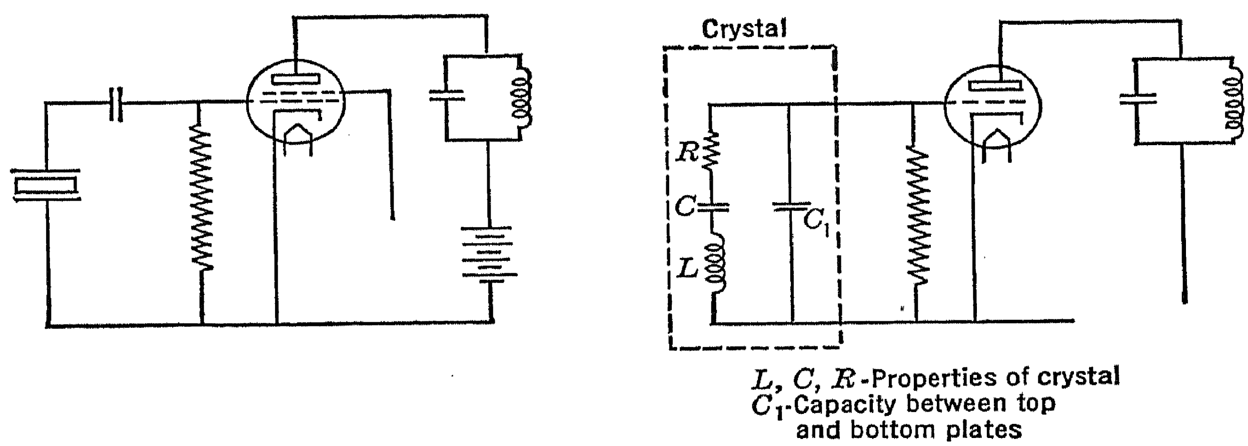


Fig. 16.5.—Crystal Oscillator.

The low damping of these crystal resonators makes them valuable as a means of controlling the frequency of vacuum-tube oscillators. Such an oscillator, together with its electrical equivalent circuit, is shown in Fig. 16.5. It will be seen that this oscillator is similar to the conventional tuned plate, tuned grid circuit. Many other circuits may be used, but this type is perhaps the most common.

By using a very thin crystal and establishing vibrations normal to the crystal face, crystal-controlled oscillators can be made to operate at frequencies up to 10 megacycles without difficulty, and frequencies as high as 20 megacycles are possible. However, even this upper frequency is not great enough for the carrier required by television.

To obtain frequencies of 40 to 100 megacycles or higher with a crystal oscillator it is necessary to multiply the frequency of the output by a factor of four or more. This can easily be done by means of a harmonic generator, a multivibrator synchronized at a multiple of the oscillator frequency, or some other form of frequency multiplier. Of these frequency multipliers the harmonic generator is perhaps the most common. The

harmonic generator consists essentially of a class C amplifier whose plate tank circuit is resonated to the desired harmonic of the input. The curves shown in Fig. 16.6*b* are given to clarify its operation. The bottom curve represents the voltage input supplied from the crystal oscillator to the grid of the generator. The plate current and load voltage wave forms are also indicated. Fig. 16.6*a* is the schematic circuit diagram of such a generator. Because the current output of a tube operated class C is high in harmonic content, such a device is a fairly efficient power generator. Actually the power delivered by this type of generator is

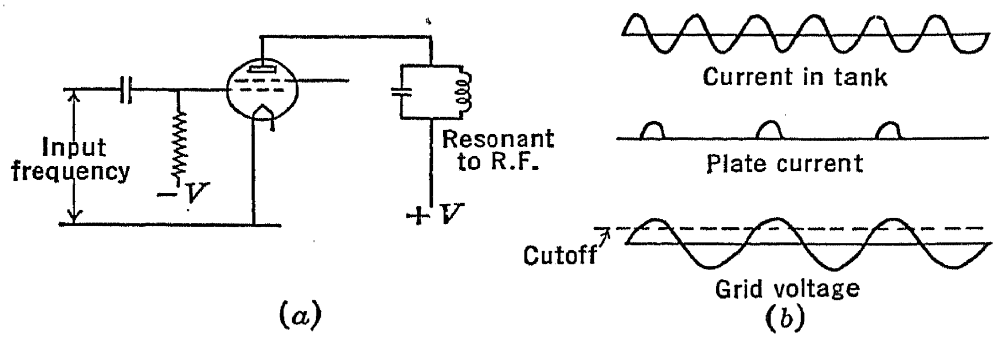


FIG. 16.6.—Schematic Diagram of Harmonic Generator and Its Operation.

roughly inversely proportional to the order of harmonic used; thus, operating on the second harmonic, it will deliver about 60 per cent of the power that the same tube would deliver as an ordinary class C amplifier, while on the fourth harmonic the power would be 30 per cent.

A quartz crystal oscillator can deliver as much as 50 watts of power output. However, at very high frequencies it is advantageous from the standpoint of stability to make the power required from the oscillator a minimum. This avoids overheating the crystal as a result of internal fric-

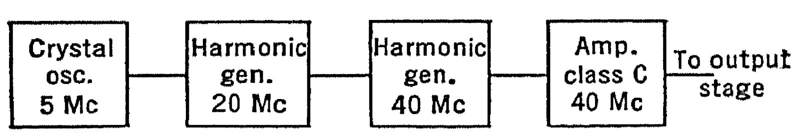


FIG. 16.7.—Block Diagram of Carrier Generator.

tion. The power required to operate the harmonic generator is small and therefore usually no amplification is required between it and the crystal oscillator. A complete crystal-controlled carrier generator is shown as a block diagram in Fig. 16.7.

The low-loss "transmission line" oscillator has also proved a satisfactory master source for the carrier power generator. This oscillator is a vacuum-tube oscillator controlled by a resonant circuit. The resonant circuit is not a conventional coil and condenser tuned circuit, but instead a quarter-wavelength concentric transmission line. This type of resonant

circuit is practical only at very high frequencies, where a quarter-wave line becomes a matter of a few feet. Such a concentric line tank circuit consists of two cylinders, one inside the other, whose length is equal to a quarter of the wavelength of the oscillations involved; i.e., for 40 megacycles the length is slightly over 6 feet. The two cylinders are connected together at one end by means of a heavy metal disk, which also serves to hold the cylinders rigidly concentric. The surface conductivity of the tank circuit must be very high because the "skin effect" at ultra-high frequencies is so great that all the current is localized within a fraction of a mil of the surface. It is particularly important that the contact at the bridging disk be very good; therefore, often not only is solder or brazing used to hold the disk in place, but also the joint, and preferably the whole tank circuit, is plated with copper or silver.

The stability of this type of tank is a function of the Q of the circuit, where Q has its ordinary meaning of $2\pi fL/R$. Ordinarily the frequency drift Δf due to normal changes in driving power, load, etc., will not be greater than that which causes a 30 per cent decrease in the impedance of the tank. On this basis it can be shown that:

$$\frac{\Delta f}{f} = \frac{1}{Q},$$

where f is the resonant frequency. The Q of the circuit is related to the geometry, physical size, and material of the line. It will be shown in a succeeding section that a diameter ratio of 3.6 between inner and outer conductor yields the greatest efficiency and consequently the highest Q for practical construction. Inasmuch as the factor Q is approximately proportional to the circumference of the conductors (because surface rather than volume determines the conductivity), it is desirable to make the diameter of the outer cylinder large. In practice, diameters of 1 foot or $1\frac{1}{2}$ feet are employed. When properly constructed a concentric line tank circuit can be made to have values of Q between 5000 and 10,000 and have a frequency stability of a few hundred cycles in 40 million.

The oscillator is frequently a conventional tuned plate, tuned grid type. For various practical reasons, the concentric line is usually placed in the grid circuit, while the tuned plate load consists simply of a relatively low-Q two-wire transmission line. Fig. 16.8 shows schematically a typical push-pull oscillator using a high-Q tank. The driving unit in the example illustrated is loosely coupled to the tank circuit by a short loop near the inner conductor close to the point where it is attached to the outer cylinder. This loop is, in reality, a transformer and links the magnetic flux produced by the current in the inner cylinder. The location close to the bridging disk is, of course, chosen so as to be near the

current maximum of the standing wave which exists on the line. The plate tank is coupled to the load.

Obviously, if a high Q is to be maintained, the load on the concentric tank must be small. Moreover, the load must remain relatively constant

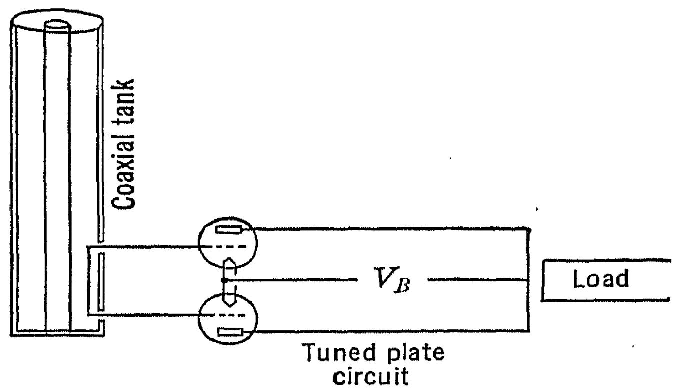


Fig. 16.8.—Push-Pull Concentric Tank Oscillator.

For this reason the coupling between the carrier amplifier and the oscillator tank circuit is made very loose so that the carrier voltage applied to the grid of the succeeding tube is small, even though this makes necessary an additional stage of carrier amplification.

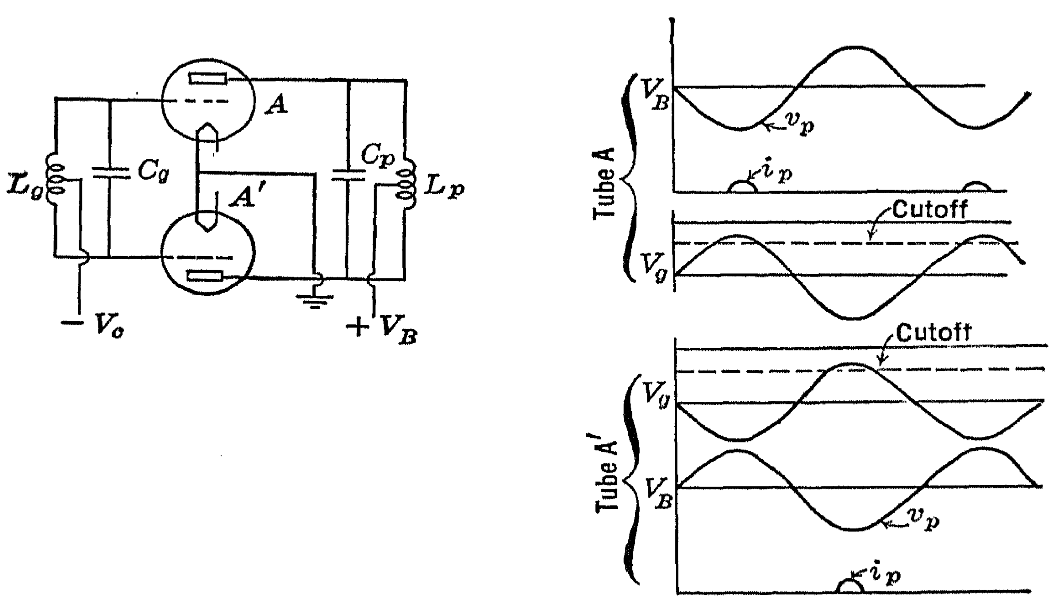


Fig. 16.9.—Circuit and Operation of Class C Push-Pull Amplifier.

It might be said that the conditions for frequency precision in this type of oscillator are exactly those required by an accurate clock. The energy in the oscillating circuit or member must be large compared to the energy added to and withdrawn during any one cycle. This is possible only if resistance and load losses are small, and the mass or inductance of the oscillating element is large.

The ultra-high frequency from the master oscillator, whether it is crystal controlled or controlled with a concentric tank, must be greatly amplified before it has the power level required by the output stages. Class C amplifiers are frequently employed for this purpose. A class C push-pull amplifier is shown in Fig. 16.9, together with the form of the voltage wave applied to the grid and plate of the tubes and of the plate current. The high efficiency of this form of amplifier is a consequence of the fact that the tubes are biased to cutoff except when the plate voltage across the tubes is a minimum, making the ratio of power dissipated in the tube to power delivered to the load a minimum. Efficiencies of 85 per cent or more can be obtained with such amplifiers, where the plate load impedance is a sharply resonant circuit with little damping, i.e., a high-Q circuit. This is to be compared with the 70 per cent efficiency of a class B amplifier, where each tube is biased so that it delivers current for half the radio-frequency cycle, and the very much lower efficiency of the conventional class A circuit, typified by the single-ended amplifier discussed in Chapter 14.

16.3. Modulation. Having considered the generation of carrier power, and in a preceding chapter the amplification of the video signal, the next problem to be examined is that of modulating the former by the latter. This is one of the most difficult problems facing the television engineer.

Modulation, as has already been pointed out, consists of varying the amplitude of the carrier in such a way that the envelope of the carrier wave corresponds to the modulating signal. The principal methods used at present for sound transmission are based on an amplifier stage whose gain is proportional to the modulating signal. Fundamentally, these may be divided into amplifiers whose gain is varied by a control of the plate voltage, and those whose gain is controlled by the mean value of the grid potential. These two classes are themselves subdivided into specific circuits of varying degrees of complexity, designed to fulfill the requirements of particular types of transmitters.

The basic circuit of a plate-modulated stage is illustrated in Fig. 16.10. The variable gain amplifier is shown as a push-pull circuit, consisting of two power tubes A and A'. The tuned grid circuit is excited from the carrier generator, as in an ordinary radio-frequency amplifier. The plate voltage is supplied through the coupling impedance Z_c, and depends upon the voltage drop through the impedance. The current through Z_c is not only the current supplied to the amplifier tubes, but also that to the modulator tube B. When the grid of B is negative so that the plate current is small, the voltage drop through the impedance is small, and the voltage applied to the amplifier stage a maximum. Under these con-

ditions the radio-frequency power delivered to the load is high. On the other hand, when the modulator tube is conducting so that the voltage drop through the coupling impedance is large, the gain of the amplifier is low. By proper adjustment of the circuit constants involved, the modulation can be made linear.

In sound transmission, the efficiency of this type of modulator can be made fairly high where the modulator stage is a class B push-pull amplifier. This is possible only where the modulation varies symmetrically around a mean carrier level. In television broadcasting this a-c transmission is not economical, as has already been pointed out. Instead the d-c component of the video signal is included. As a consequence the variable gain stage must be controlled by a single-ended class A amplifier. This greatly decreases the efficiency of the system.

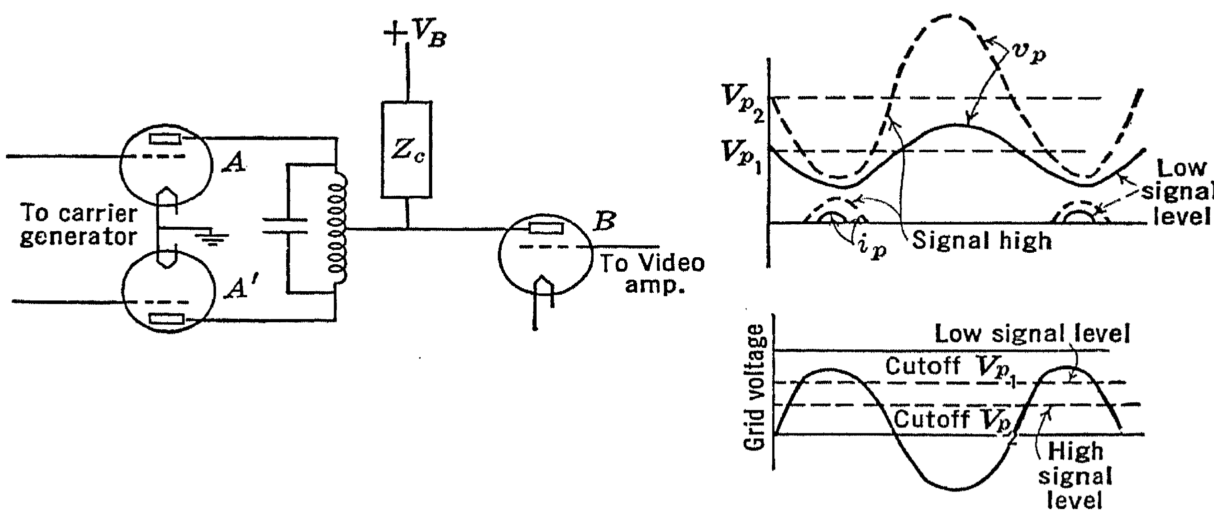


FIG. 16.10.—Schematic Circuit Diagram of Plate Modulator.

A second serious objection to this type of modulator for television transmission is the requirement placed on the coupling impedance. Like the interstage coupling of the video amplifier, it must be constant in magnitude and time delay over the video-frequency band. Furthermore, it carries the entire d-c component of the power amplifier. Because of this second condition, a pure resistance cannot be used without making the loss of power prohibitive. The choice of coupling impedance is limited to an inductance or a filter net. The former has been employed but is unsatisfactory because of the extreme range of frequencies involved. A filter net which includes the tube capacity, together with low-impedance modulator tubes, offers a possibility of improving the efficiency of this form of modulator. At present these conditions cannot be met because neither suitable tubes nor filters are available. However, much promising experimental work is in progress along these lines.

Grid modulation is somewhat better suited to television transmission,

although even this method is subject to serious limitations. A schematic diagram of a grid modulator is shown in Fig. 16.11. Like the preceding modulator it is essentially a variable gain, ultra-high-frequency power amplifier. The gain in this instance is controlled by the grid bias of the stage. Constant amplitude radio frequency is supplied to the power tubes, A and A', in phase opposition through the tuned circuit which couples them to the carrier generator. The bias for the amplifier tubes is introduced at the midpoint of the grid tank, and is obtained from a voltage supply through the coupling impedance Z_c. This impedance also carries the plate current of the modulator tube B. The mean grid voltage applied to the amplifier has its maximum positive value when the modulator tube is at cutoff, and becomes more negative as the plate current

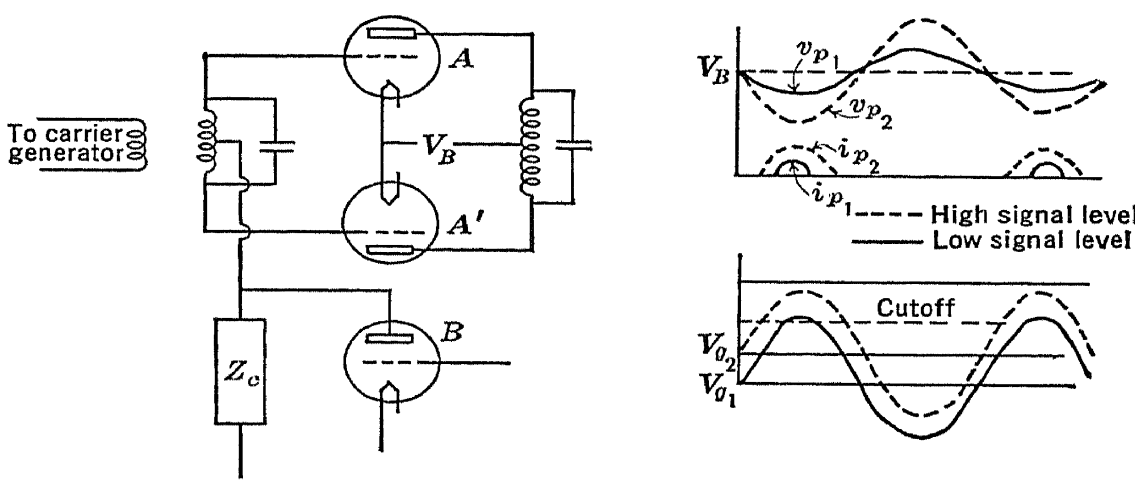


FIG. 16.11.—Schematic Circuit Diagram of Grid Modulator.

in the latter increases. In most installations the bias voltage supplied to the coupling impedance is so adjusted that when the modulator is at cutoff each amplifier passes current for about 180° of the r-f cycle. As the grids are made more negative, the angle over which the tubes are conducting decreases, and, at the same time, the peak value of their plate current becomes lower. Thus the power delivered to the load decreases.

The efficiency of the transmitter per se is of relatively little importance. However, since the efficiency of the power amplifier stage of the modulator determines the amount of useful power which can be obtained from tubes having a given rating, it warrants a more quantitative consideration.

Apart from the power dissipated in the modulator tube itself, the efficiency which can be obtained at maximum r-f output depends upon maximum plate current of the amplifier tubes, the minimum plate voltage for which it can be obtained, and the impedance of the tuned plate load. The significance of and reason for this relationship are shown in

Fig. 16.12. The first two factors are functions of the tube design and are fixed by the tubes available. The third factor depends upon the properties of tuned circuits, and its maximum value is limited by the frequency bandwidth and the capacity of the amplifier tubes.

The following discussion of the power limitations based upon the bandwidth which must be accepted by the plate tank circuit applies equally to plate or grid modulation.

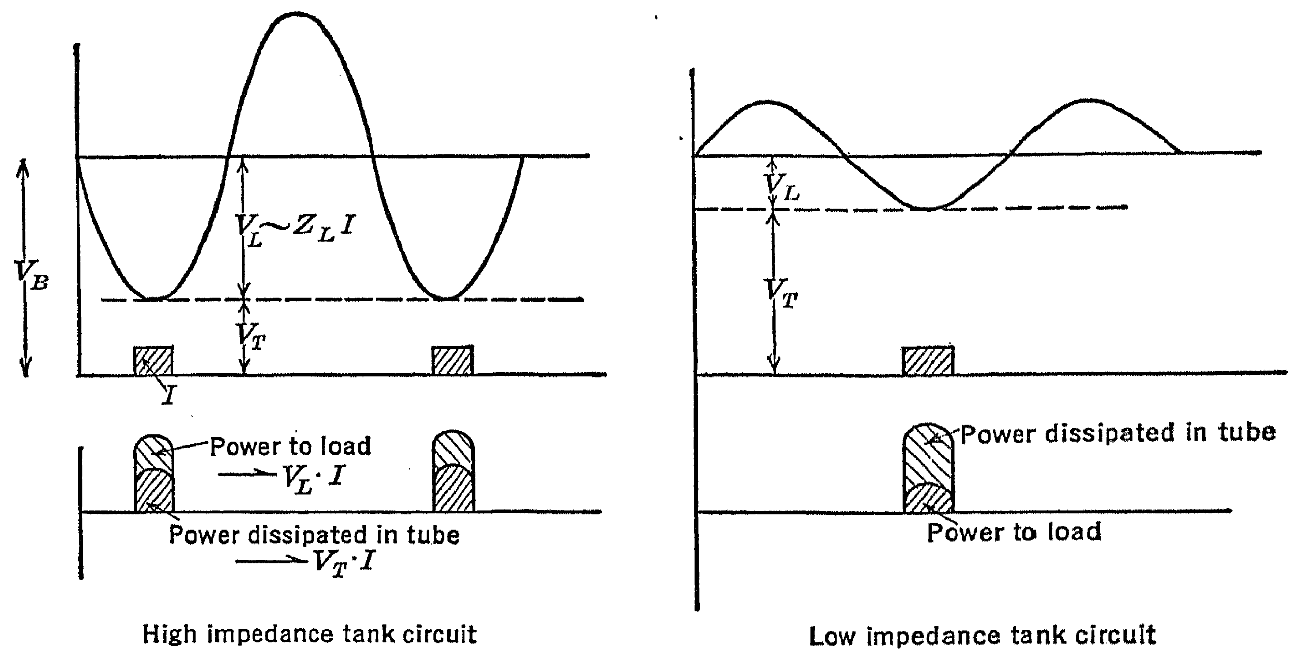


FIG. 16.12.—Power Relations in Class C Circuits.

The power W delivered to the load from a class C amplifier is given by:

$$W = \frac{(E_b - E_{\min.})I}{2}, \tag{16.1}$$

where E_b is the applied voltage across the load and tube, $E_{\min.}$ is the minimum voltage across the tube at the r-f negative peaks, and I is the peak value of fundamental a-c component of the plate current. It will be seen that the power is dependent upon the r-f voltage amplitude $E_b - E_{\min.}$ as well as upon the tube current and voltage which are determined by the amplifier characteristics. The r-f voltage is in turn proportional to the load impedance Z since

$$E_b - E_{\min.} \doteq IZ. \tag{16.2}$$

At resonance, when the current and voltage are in phase, the resonant circuit presents a purely resistive load, and the power is given by

$$W = \frac{I^2R}{2}. \tag{16.3}$$

Obviously the power increases with increasing R.

Were only a single frequency involved, the resistance of the load at resonance could be made very large by employing a high-Q circuit. However, the impedance of the circuit must remain high over the band of frequencies covered by the sidebands. If the power at the extremes of the modulation frequency f_m is not to be less than 70 per cent of that at carrier frequency, the Q of the circuit cannot be greater than that given by the relation

$$\frac{f_m}{f_c} = \frac{1}{2Q}. \tag{16.4}$$

At carrier frequency the impedance of the load is

$$R = 2\pi f_c LQ, \tag{16.5}$$

and furthermore,

$$2\pi f_c L = \frac{1}{2\pi f_c C}, \tag{16.6}$$

where L is the inductance of the load and C the total capacity. Substituting from Eq. 16.6 and Eq. 16.4 in Eq. 16.5, the maximum impedance of the load at resonance is

$$R = \frac{1}{4\pi f_m C}.$$

Consequently the power output available is

$$W = \frac{I^2}{8\pi f_m C}.$$

Therefore the power delivered depends not only upon the current and voltage characteristics of the driving tube, but also upon its capacity, since this capacity is the minimum that can appear across the load. For tubes in general use at present this capacity places a very serious limitation on the performance that can be expected from them. This is particularly true for the class of tubes designed for high power output.

From the foregoing it is apparent that any modulation system, in which there is modulation of the carrier current in a tank circuit, is subject to this limitation.

Systems employing absorption modulation, impedance transfer networks, etc., arranged so that the generator maintains a constant voltage across the tank circuit, irrespective of the power delivered to the antenna, overcome this difficulty, at least in part. An interesting example of this type of modulator is described by W. N. Parker.* The system in question

* See Parker, reference 10

makes use of the impedance inversion property of a quarter-wave line, and modulates the radio frequency in the antenna transmission line rather than attempting to modulate the power in the tank circuit of the output stage. A schematic diagram of the modulator is shown in Figs. 16.13*a* and *b*. The circuit consists of a conventional ultra-high-frequency oscillator coupled to a quarter-wave line at whose far end are two branches, one being a quarter-wave line leading to the modulating tubes, and the other the transmission line to the antenna. In the equivalent circuit diagram, Fig. 16.13*b*, the antenna is represented by the load resistance R while the modulating tubes are shown as the variable con-

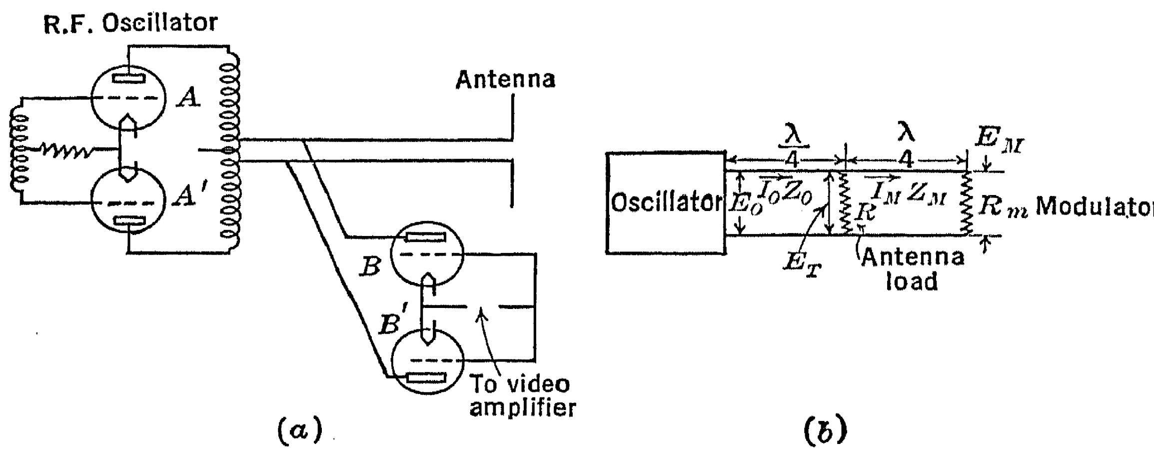


FIG. 16.13.—Parker Modulation System.

ductance element R_m. It is a property of a quarter-wave line that if its surge impedance is Z_s and the terminating impedance is R its input impedance, as will be seen in section 16.6, is given by:

$$Z_i = \frac{Z_s^2}{R}.$$

Therefore, if the modulating tubes are biased to cutoff so that R_m becomes infinite, the impedance of the line L_m at the load resistance approaches zero. Furthermore, since the line L_0 is also a quarter wavelength long, its impedance at the generator becomes very high, and no power is delivered to the load. Conversely, if the grids of the modulating tubes are made positive, the value of R_m is small, and the input impedance to the line L_m is high. Under these conditions the line L_0 is essentially terminated with the antenna load resistance R and its impedance at the generator becomes Z_0^2/R; hence power is delivered to the load.

While this modulator does not require a rapid variation of the energy in the plate tank circuit of the power amplifiers, and therefore permits a more efficient utilization of these tubes, it is not independent of frequency because of the two quarter-wave lines involved. At carrier frequency

these lines can be made equal to exactly $\lambda/4$. However, this condition will not be fulfilled at the sideband frequencies, so that for these frequencies the impedance presented by the line at its input end is not purely resistive. The capacity between plate and ground of the modulator tubes and the capacity of the line itself have an effect on the response of the line which is quite analogous to that of the parallel capacity of a tuned circuit. This limits the amount of power that can be modulated with given tubes when a wide frequency band is involved. Furthermore, unless the effective resistance of the modulator tubes is very low when their grids are made positive, the impedance of the line L_m at the antenna load

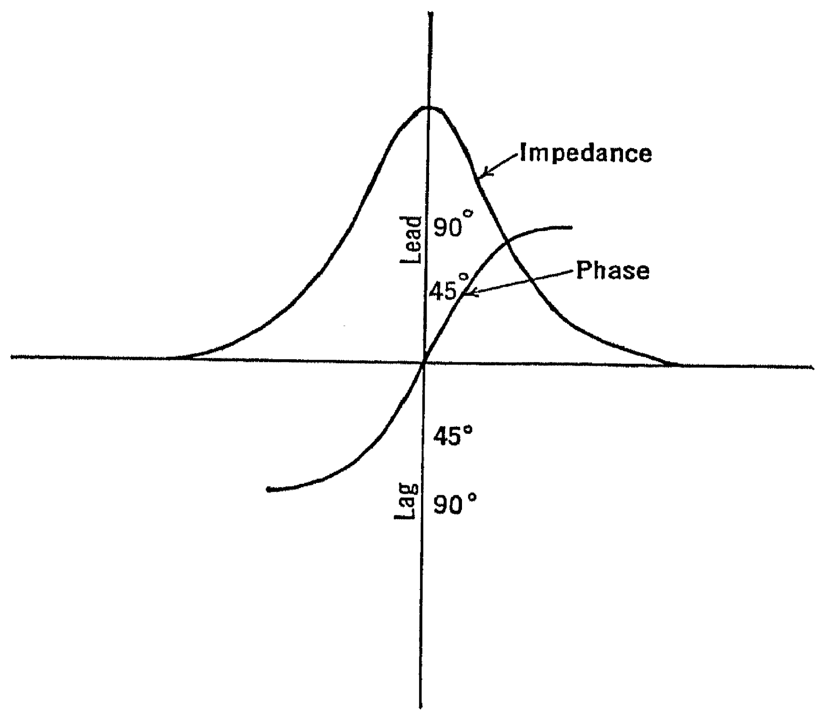


FIG. 16.14.—Impedance Characteristics of Tuned Circuit.

R will not be high enough to prevent a large part of the power which should go to the antenna from being dissipated in the modulator tubes. It will be seen, therefore, that the tube requirements for this modulator are identical with those of the grid and plate modulators, namely, a low output capacity and a low minimum plate resistance. Low-power tubes are available which meet these requirements fairly satisfactorily, but high-power tubes have not yet been made to operate efficiently.

The possibility of modulating at low power level for single sideband operation and amplifying the modulated signal up to the desired power has already been mentioned. On the surface this appears to be a very attractive solution to the modulation problem, because it cuts to half the bandwidth which the power amplifiers must handle. A difficulty is found, however, in the fact that a resonant circuit must be used as the plate load

if efficiency is to be obtained from an r-f amplifier. The impedance over the resonance peak goes from capacitative reactance, through pure resistance, to inductive reactance, as shown in Fig. 16.14. Since for single sideband transmission the carrier lies at one or the other extreme of the band to be passed, the plate load for carrier frequency will be complex, and the load line an ellipse. Therefore the carrier amplification will be inefficient. On the other hand, if the resonant circuit is made broad enough so that the carrier can be placed at the center of the peak, making the load at the carrier frequency purely resistive, then, from the same considerations given above, the plate load impedance must necessarily be low and the amplification inefficient.

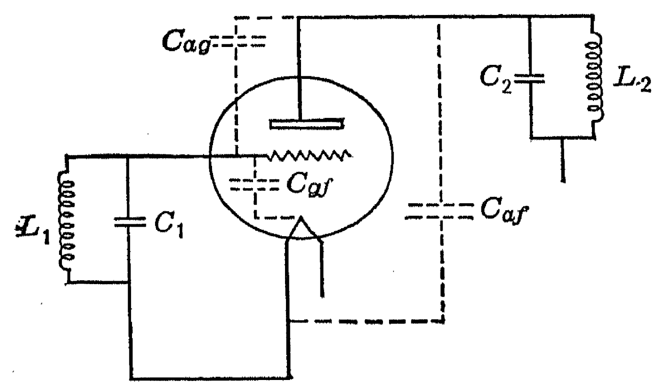


FIG. 16.15.—Interelectrode Capacities of a Triode.

16.4. Neutralization. The geometry of any grid-controlled vacuum tube is such that the various elements have a certain capacity to one another. Because of this capacity, when a voltage is applied to the anode a voltage will be induced on the grid, resulting in a form of feedback. If an alternating voltage is applied to the anode, the importance of this feedback becomes rapidly greater as the frequency is increased. Thus at radio frequency it is necessary to take corrective measures to overcome the effects of this feedback, while at ultra-high frequency this correction is a major problem.

The principal interelement capacities in a triode are indicated in Fig. 16.15. From this diagram it is evident that, in the absence of the external circuits, the capacities C_{ag} and C_{gf} are, in effect, a voltage divider between grid and anode, and the induced voltage on the grid will be in phase with the anode voltage. The impedance of this voltage divider is inversely proportional to the frequency.

A purely resistive plate load causes the plate voltage to vary in such a way that the potential induced on the grid varies in phase opposition with the applied grid excitation. However, if the external grid and plate circuits are networks containing inductance and capacity, for example,

tuned circuits, the condition of phase opposition is no longer necessarily true. Under these conditions, the circuit as a whole may become unstable and oscillate. Without presenting an analysis of the circuit, this can be seen from the fact that at a frequency close to the resonance of the tuned circuits they may adjust their relative phase in such a way that the effective circuit is essentially a conventional oscillator circuit.

In order to stabilize an ultra-high-frequency amplifier, it is necessary to arrange the external circuit in such a way that a voltage is applied to the grid opposing that supplied by the interelectrode capacity. The proc-

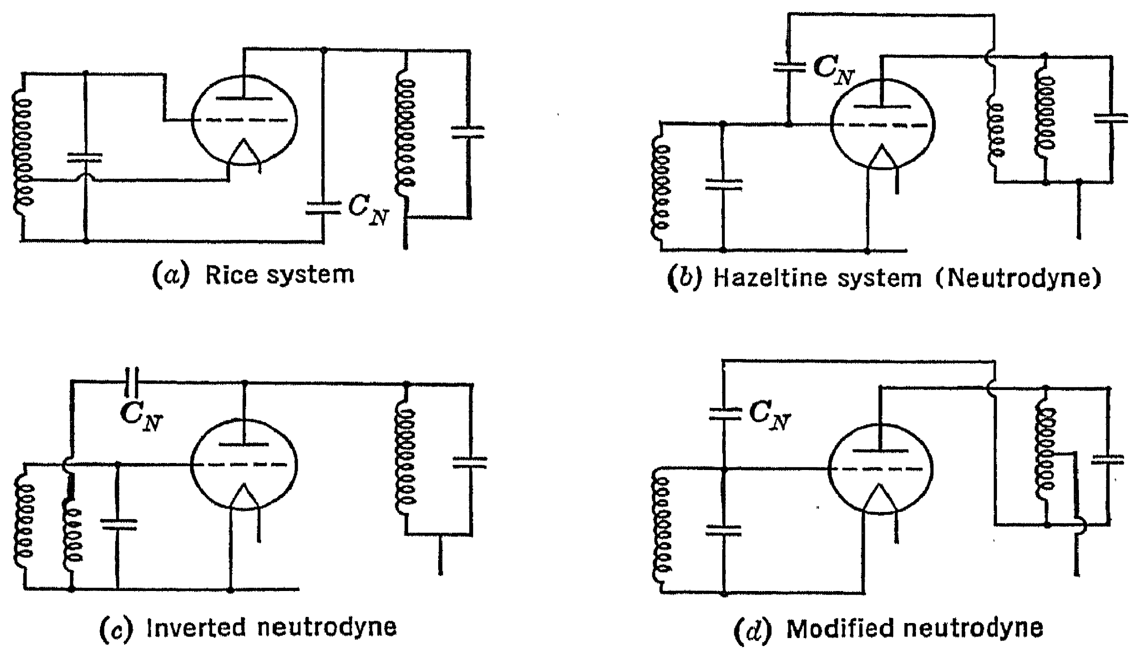
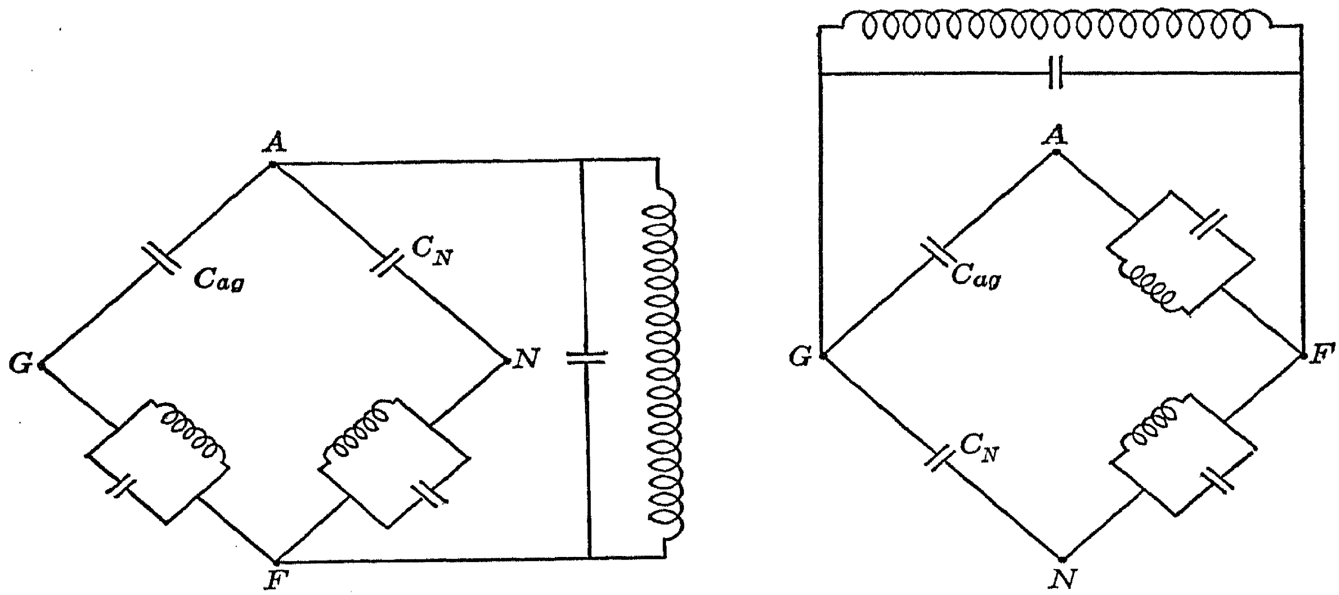


Fig. 16.16.—Systems of Neutralizing Unwanted Capacities.

ess of compensating for the unwanted capacities is known as neutralizing. There are a great many ways of carrying out neutralization, a few being illustrated in Fig. 16.16.

The action of these circuits is more easily visualized if they are redrawn in bridge form as shown in Fig. 16.17. Circuits of the type shown in Fig. 16.17*a* are known as grid neutralized, and correspond to the actual connections of Fig. 16.16*a* and *c*, while those of Fig. 16.17*b*, corresponding to *b* and *d* in Fig. 16.16, are plate neutralized.

For very high-frequency work experience has shown that bridges of the type illustrated, wherein the circuit elements used in symmetrical arms are physically different, can be balanced only with extreme difficulty. Therefore it is general practice to use push-pull stages, making it possible to design a bridge which is physically symmetrical. A neutralized push-pull amplifier is illustrated in Fig. 16.18, together with the corresponding bridge.



(*a*) Grid neutralized (*b*) Plate neutralized

FIG. 16.17.—Unsymmetrical Neutralizing Bridges.

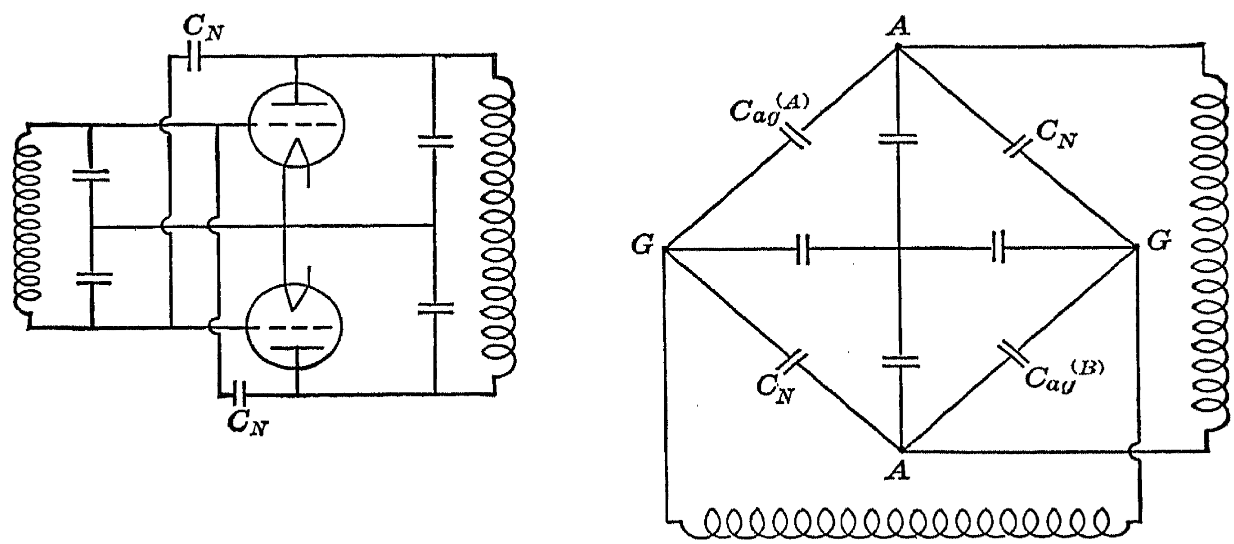


FIG. 16.18.—Push-Pull Neutralizing Circuit and Equivalent Bridge.

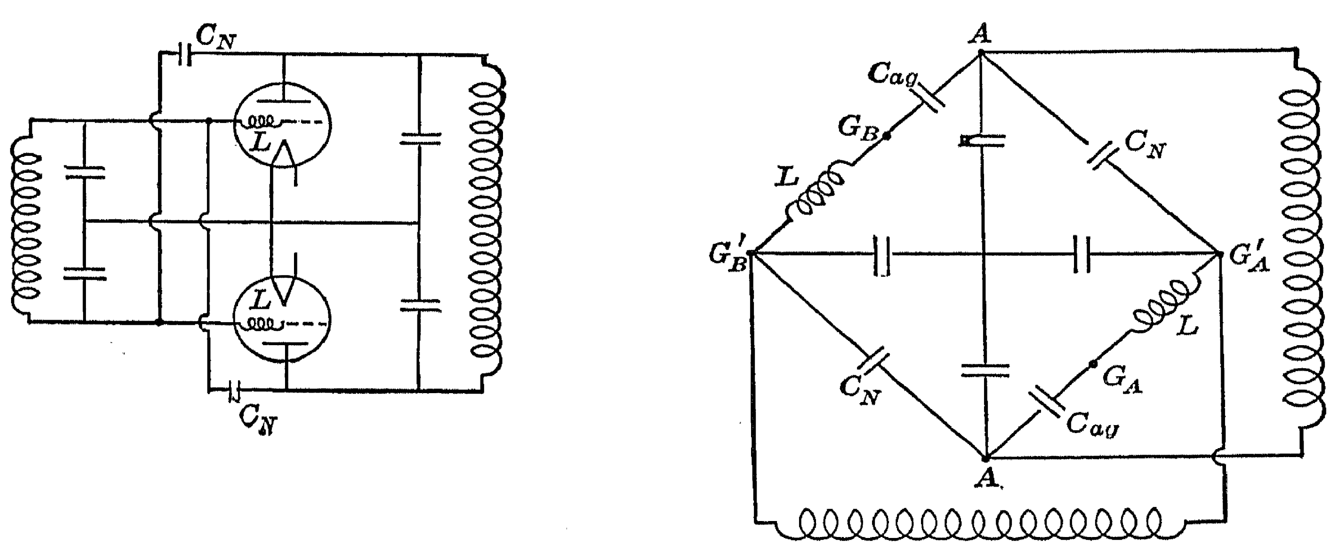


FIG. 16.19.—Effective Grid Lead Inductances in Neutralizing Circuits.

At the ultra-high frequencies, unlike intermediate and high frequencies, it does not suffice to compensate for the interelectrode capacities alone. In addition, it is necessary to take into account the inductance of the leads, which, as is evident from Fig. 16.19, makes it difficult to apply neutralizing voltages to the points required for full compensation. For simplicity, only the inductance of the grid lead is shown in the figure.

It is obvious from this figure that, if the circuit elements of the bridge are such that $C_{ag} = C_N$, and a voltage is applied to the anode points, a voltage will appear at the true grid points G_A and G_B, and consequently between the filament points and the grids. In order to illustrate the magnitude of the voltage thus induced, the following typical tube characteristics will serve for a quantitative example:

f_c = 50 megacycles, carrier;
V_{af} = 3000 volts anode swing;
C_{ag} = 20-$\mu\mu$f grid anode capacity;
L_g = 0.15 μh inductance of grid lead.

The voltage V_g appearing on the grid will be:

$$V_g = V_{af}\, 4\pi^2 f_c^2 C_{ag} L_g = 900 \text{ volts,}$$

which voltage is to be compared with the 1000 to 1500 volts ordinarily used to drive such a tube.

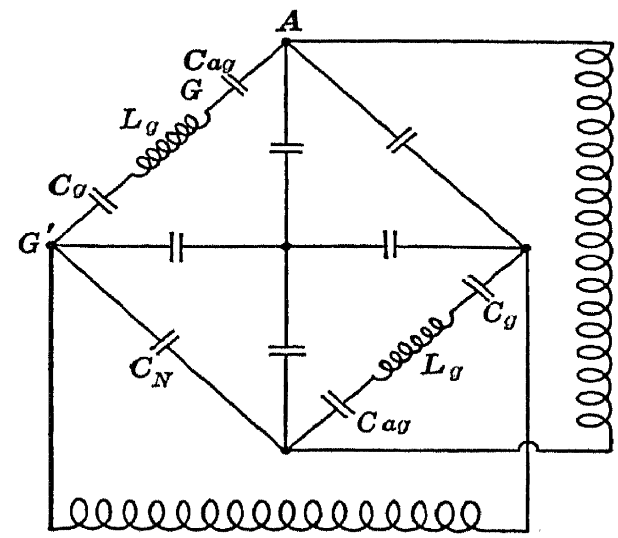


FIG. 16.20.—Compensation for Grid Lead Inductances in Neutralizing Bridge.

This bridge can be neutralized in such a way as to overcome the effect of the grid lead inductance if a capacitance is put in series with the grid lead, such that its impedance just equals the impedance of the lead.* This is shown diagrammatically in Fig. 16.20. The conditions for balance and no voltage at the true grid points are:

$$C_N = C_{ag}$$
$$\frac{1}{2\pi f_c C_g} = 2\pi f_c L_g. \tag{16.7}$$

It is immediately evident, however, that these conditions can be fulfilled only at one frequency. Where a wide communication band is used, as in television, the compensation will not be complete for the sideband frequencies. Continuing the illustrative example, assuming a bandwidth Δf

* See Davis and Green, reference 9.

of 4 megacycles either side of the carrier, the added grid voltage at the sideband extremes is calculated as follows:

$$V_g = 4\pi^2 V C_{ag} L_g (f_c^2 - [f_c \pm \Delta f]^2) \tag{16.8}$$
$$\cong 8\pi^2 V C_{ag} L_g \Delta f \cdot f_c \tag{16.8a}$$
$$\cong 145 \text{ volts.}$$

This voltage is in phase at frequencies above the carrier, and has phase opposition at frequencies below the carrier, therefore resulting in a slight

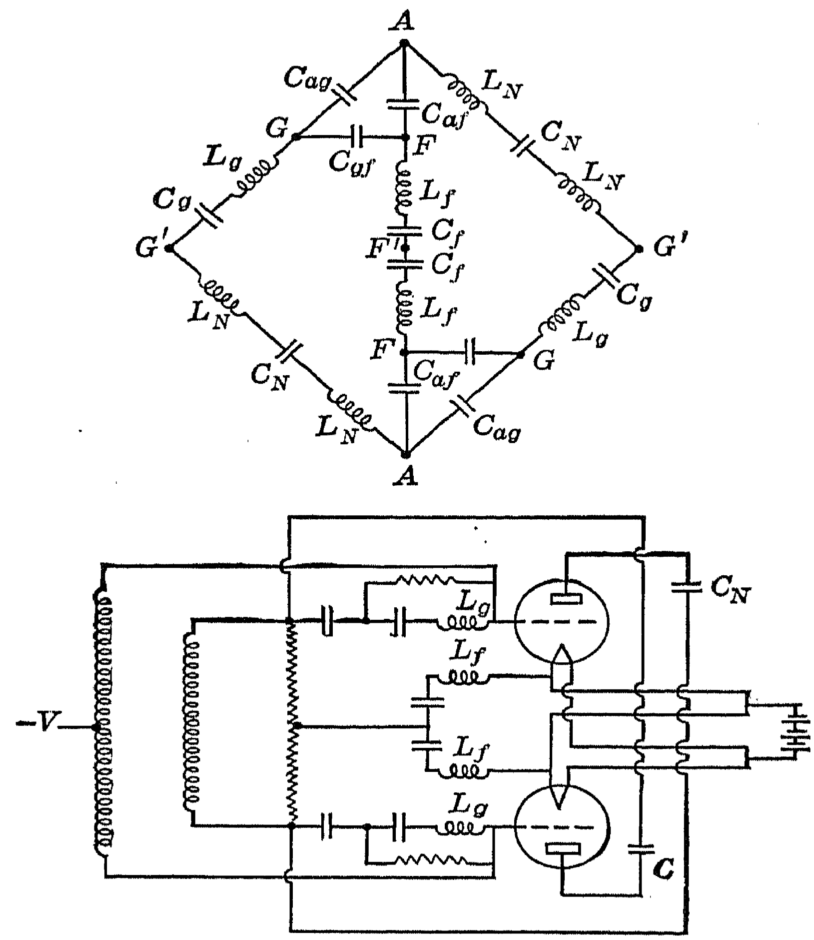


FIG. 16.21.—Neutralizing System Utilized in Philips Transmitter PCJ.

dissymmetry of the sidebands. However, for most practical transmitters this degree of neutralization is adequate.

The simplification was made of omitting the lead effects of all tube electrodes except the grid. Actually this is not justifiable, and the inductance of all the leads must be taken into account in constructing the neutralization bridge. Fig. 16.21 illustrates a complex bridge which includes compensation for filament and neutralizing condenser leads, as well as the grid lead. The amplifier circuit which is represented by this bridge is shown in Fig. 16.21. This type of circuit is employed in the output stages of transmitter PCJ operated by the N. V. Philips Co.*

* See Nordlohne, reference 11.

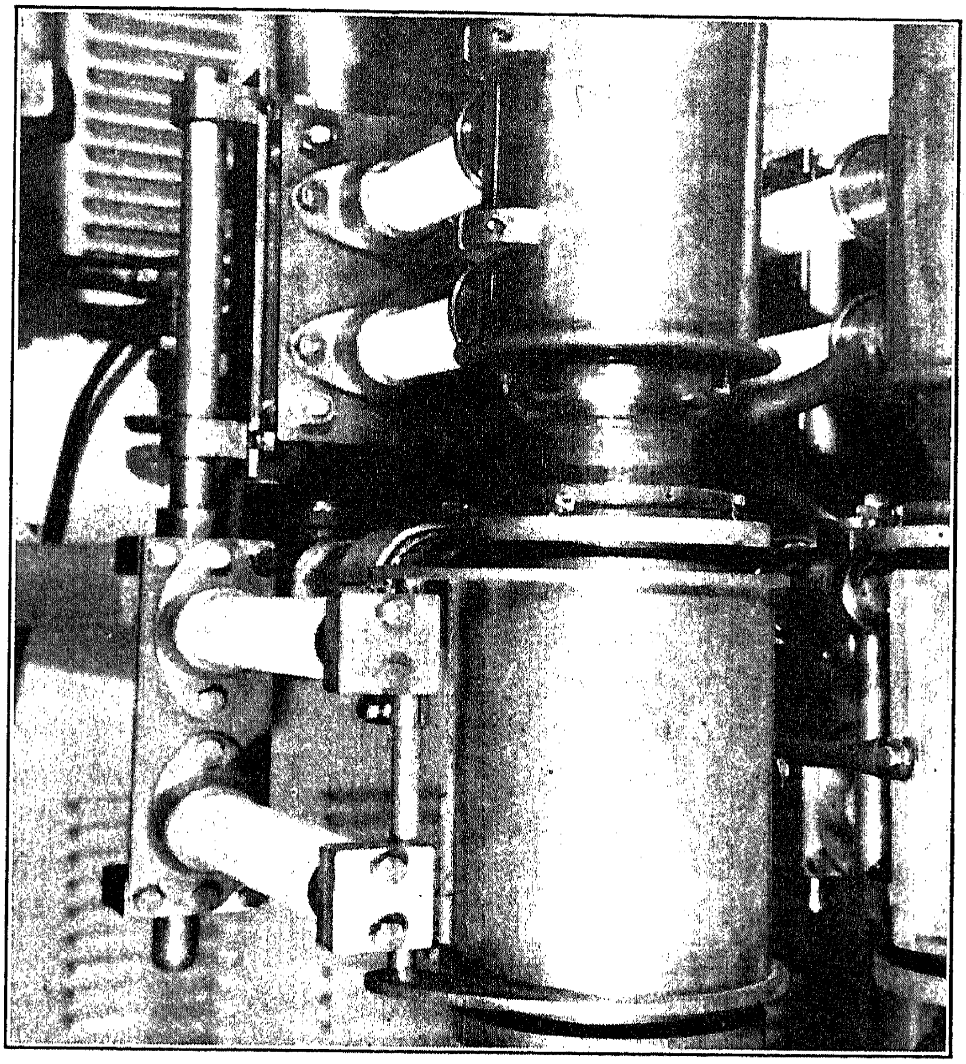


FIG. 16.22.—Ultra-High-Frequency Transmitting Tube with Sleeve-Type Neutralizing Condenser.

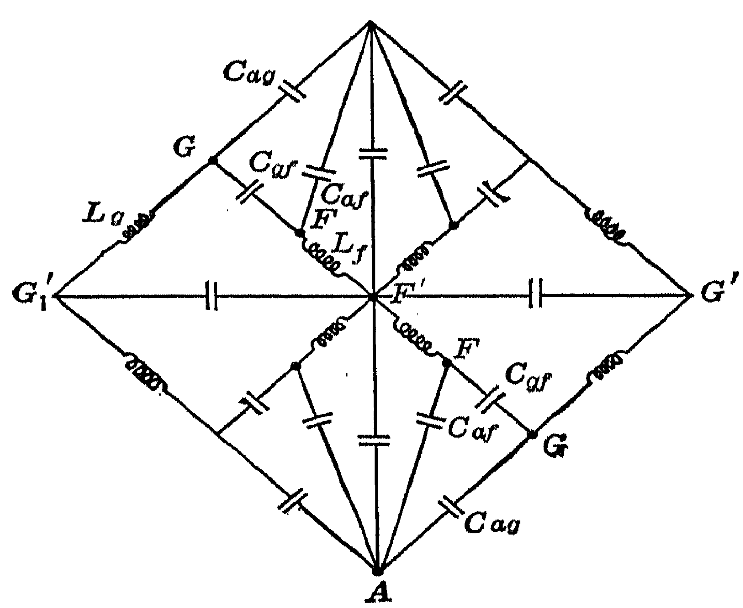


FIG. 16.23.—Bridge for Frequency-Independent Neutralization.

As in the previously discussed example, this circuit is neutralized for one frequency only.

In order to minimize the frequency dependence noted in connection with the last two neutralizing systems, it is necessary to apply the compensating voltage to the tube elements themselves, rather than to the leads. Special short-wave tubes are made to fulfill this condition fairly well. It is accomplished by using the electrode (more specifically the grid) as one plate of a neutralizing condenser (C_N in Fig. 16.18) and a sleeve fitting over the tube as the second plate. A tube and condenser of this type is shown in Fig. 16.22. A bridge for complete frequency-independent neutralization is diagrammed in Fig. 16.23. This bridge has very high symmetry and results in almost exact compensation at all frequencies, the only unbalance which can exist being due to voltage variations along the electrodes within the tube, which effect, at television frequencies, is very small.

16.5. Ultra-High-Frequency Power Tubes. The difficulties encountered in designing tubes to operate at ultra-high frequency increase very rapidly with the power output required. This, of course, is because of the contradictory demands on the construction. In order to dissipate large amounts of power and to withstand high voltages, the physical size must be large. On the other hand, the dimensions must be small enough so that the elements are only a small part of a wavelength, so that the interelectrode capacities are small, and so that the transit time of the electrons will be short. A compromise must be effected between these conflicting requirements.

Certain specific characteristics made evident from the discussion in the foregoing sections may be listed as follows:

1. The output d-c resistance and capacity must be as low as possible (i.e., I_p/E_p large).
2. The tube must be capable of handling the required voltages, and of dissipating the power incident on the tube losses.
3. The input admittance must be low.
4. The geometrical configuration must be such that the desired potentials may be applied to the tube elements.

Each of these general requirements merits consideration, from the standpoint both of operation and of construction.

The first requirement follows directly from the discussion in section 16.3, in which the question of the output circuit of a broad-band modulator or amplifier is considered. Here it is shown that the load impedance is inversely proportional to plate capacity. Hence, since the carrier voltage for a given voltage on the tube during the time it is passing current

depends upon the load impedance, it is inversely proportional to the capacity. Furthermore, the larger the current which can be driven through the tube for a given applied voltage, the greater the power that can be delivered to the load.

From the point of view of construction, the first item means that the area of the elements must be small and the spacing as large as possible in order to reduce the capacity. These conditions must be fulfilled without too great a reduction in available plate current. Therefore, emission from the cathode must be great and the elements so designed and disposed that a large electron current can reach the anode with a minimum applied voltage.

A tube operating under the conditions described for grid modulation (or as a broad-band amplifier) develops a very high voltage between anode and cathode during that part of the r-f cycle when the tube is not passing current. The peak of this voltage will, of course, be $2V_0 - V_a$, where V_0 is the d-c voltage applied to the plate circuit, and V_a is the minimum voltage appearing between plate and ground. As the efficiency of the tube is increased this voltage becomes higher, since in the limit the efficiency as a function of voltage is $(V_0 - V_a)/V_0$, assuming the operating angle to be very small. Therefore an efficient power tube must be insulated to withstand high voltages both with respect to breakdown and dielectric loss. The elements must be so formed that there are no unshielded sharp points or edges from which cold discharge may take place or about which, when they are imbedded in a dielectric, an intense field may develop.

The power lost in the tube must, of course, be dissipated. This must take place at temperatures low enough not to melt or warp metal or glass parts. Also the temperature of the electrodes must be kept below that at which there is appreciable thermionic emission. Because of the small size of the electrodes, special provision (usually in the form of water cooling) must be made to carry away the heat developed by: (1) the electron current in the tube, (2) dielectric losses, and (3) the cathode heater.

From the standpoint of operation the requirement that the grid admittance be low needs almost no comment. The reactive component must be high since it limits the impedance of the external circuit which can be used when the bandwidth is great. Obviously the grid conduction must be low since the power required to drive the grid is proportional to it; in fact, if the grid conductance becomes equal to the transconductance the tube can no longer function as a voltage amplifier.

Grid capacity can be reduced only by making the tube elements small. However, this alone does not suffice. Even though the capacity of the

tube may be nearly zero when the tube is cold, it may have considerable capacity due to electron transit time under operating conditions. This effect is exactly the same as that encountered in ultra-high-frequency receiving tubes, which are discussed in Chapter 17.

The phase relation between grid voltage and current shifts with increasing frequency, owing to electron transit time, causing the existence of a real component in the displacement current. However, because of the very much higher voltages used in connection with transmitting tubes, transit time loading of the grid circuit at television broadcast frequencies is relatively unimportant. Another cause of grid loading is the fact that it is often necessary to drive the grid positive in order to make full use of the emission current from the cathode, and under these conditions the grid may draw large electron currents.

The final requirement, that of being able to apply the desired potentials to the elements, and only those potentials, is both interesting and difficult to meet. It is this requirement, or rather the inability to satisfy this requirement, that makes necessary the elaborate neutralizing systems described in the preceding section.

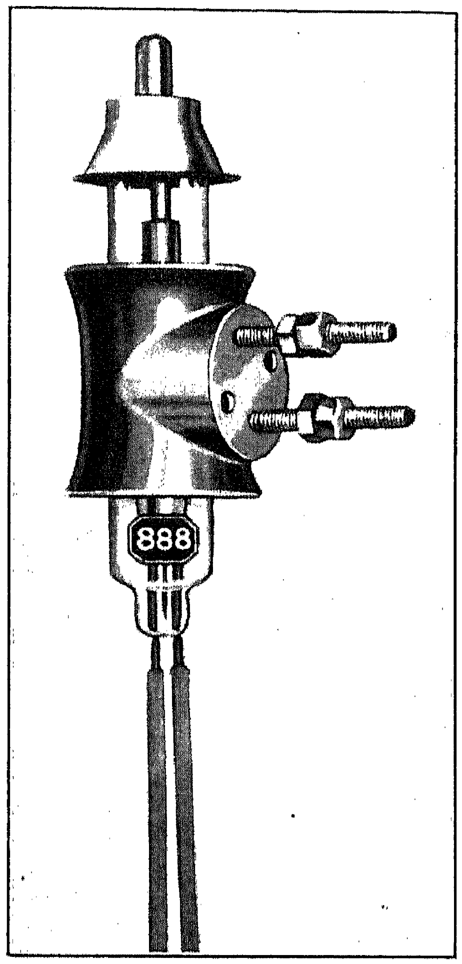
The ideal tube should be of such a size that the elements are only a negligible fraction of the wavelength used, and there should be no lumped capacitance or inductance in the tube or leads. Obviously this ideal cannot be even approximately realized. However, capacities, and in particular that between anode and grid, can be made small; leads can be made short to avoid inductance; and the tube elements so disposed as to make them as nearly as possible part of the transmission lines comprising the external circuit.

Plate to grid capacity, which is perhaps the worst offender from the standpoint of inducing unwanted potentials on the tube elements, cannot be avoided without the aid of a screen grid. Screens have been successfully used in low-power transmitting tubes, but as yet such design difficulties as cooling the screen and avoiding excessive power loss at the screen have prevented high-power ultra-short-wave screen-grid tubes from being made commercially available. Until such time as they become obtainable, neutralization must be resorted to, and the design of the tube must be such that neutralizing can be applied as directly as possible to the effective grid. This is often accomplished, as has already been mentioned, by arranging the geometry of the tube so that a sleeve can be placed outside of the tube over an exposed portion of the grid, the capacity between the sleeve and grid serving as the neutralizing condenser, thus avoiding most of the grid lead inductance.

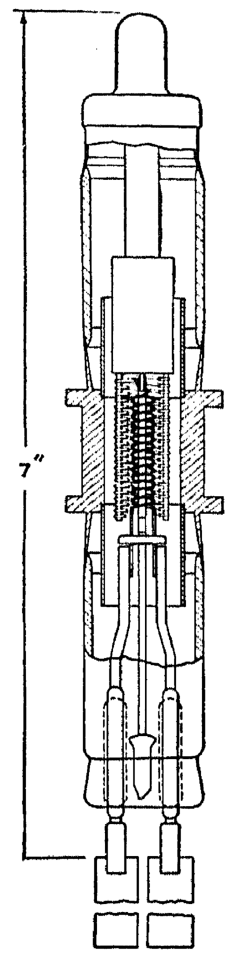
In order to illustrate the application of these design principles, the

construction of a typical intermediate-power ultra-high-frequency transmitting tube, the RCA 888, serves as an excellent example.*

The tube is shown in Fig. 16.24*a* and *b*, viewed both externally and in longitudinal section. Although the tube has a rated power dissipation of a kilowatt, its overall length (including grid and plate leads) is only 7 inches, and its diameter, exclusive of the external shields, is less than an



(*a*)



(*b*)

Fig. 16.24.—One-Kilowatt Power Tube RCA 888.

inch. When it is operated at wavelengths even as short as 1.25 meters, the total length of the grid and grid lead is less than 1/10 of a wavelength, and similarly the overall length of the filament to the external point of connection is also less than 1/10 of a wavelength. The plate itself is so arranged that it can be made essentially part of the external circuit, thus avoiding the lead problem for this electrode.

The plate is only 1 inch long and is less than an inch in diameter. It is in the form of a copper cylinder sealed into the glass insulating ends of the tube with typical tapered seals. A water-cooling jacket fits over the plate cylinder and is soldered to it, so that it can be considered electri-

* See Wagener, reference 13.

cally an integral part of the tube. The external plate circuit bolts directly to this water jacket. The jacket has flared ends which extend well past the glass-to-metal seal of the plate and act as an external shield. Internal shielding is afforded by sleeves attached to each end of the plate. These shields are necessary to avoid the intense voltage gradient that would otherwise be built up at the sharp edges of the seal.

The cathode is a pure tungsten filament supported by an internal rod attached to the cathode end of the tube, so that no internal insulating supports are required. The grid, which is made of tantalum, is likewise self-supporting. Both cathode and grid leads are of large diameter to

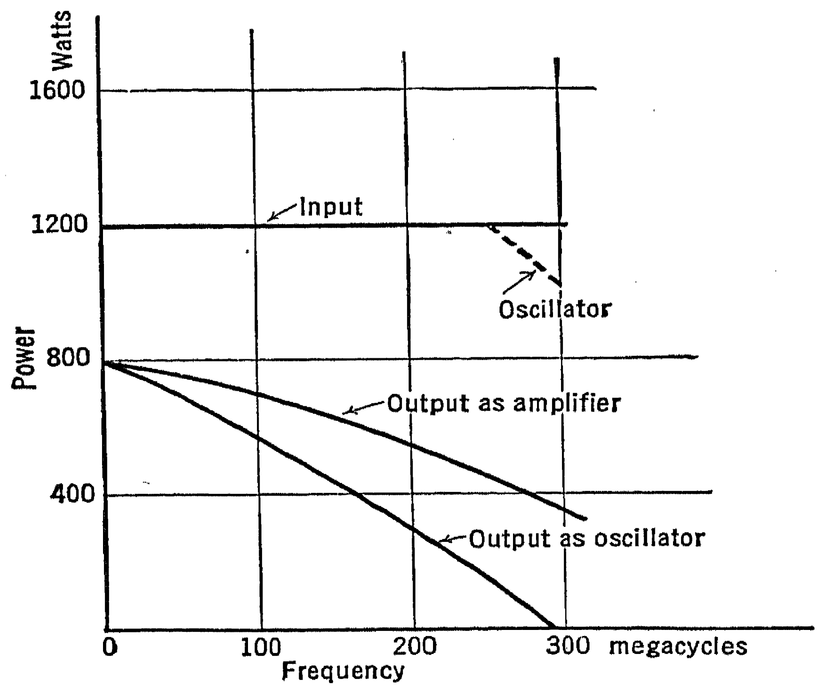


Fig. 16.25.—Amplifier and Oscillator Characteristics of the RCA 888.

reduce their inductance and to make the ohmic loss of the heavy charging currents as small as possible.

In order to minimize transit time effects, the spacing between the electrodes is made very small. Furthermore, the active portion of the tube is almost completely shielded from the glass insulation, so that no high-velocity electrons can reach these surfaces. This is quite necessary in order to avoid building up destructive gradients along the glass, or overheating it by electron bombardment.

The performance of such a tube, both as an oscillator and as a power amplifier, is shown in Fig. 16.25. When used as an oscillator this tube is of service to wavelengths nearly down to 1 meter. While amplifier tests have not been carried below 1.5 meters, they should continue to function down to wavelengths considerably below 1 meter. The operating ratings of the RCA 888 for class C operation are as follows:

Filament voltage	11.0 volts
Filament current	24.0 amperes
Amplification factor	30
Direct plate voltage	3000 volts max.
Direct grid voltage	−500 volts max.
Direct plate current	0.4 ampere max.
Direct grid current	0.1 ampere max.
Plate input	1.2 kilowatts
Plate dissipation	1.0 kilowatt

Tube type RCA 887 has a similar construction and rating except that the amplification factor is 10.

The tubes just described have only medium power rating and would be of service only in a relatively small television installation. Larger installations, such as might be used to service a large area, where 5 or 10 kilowatts carrier power is needed, require larger tubes. Fig. 16.26 illustrates larger ultra-short-wave tubes. One tube shown is an RCA 899 and has a rated dissipation of 30 kilowatts. The structure is similar to that described above, but, because of its greater power capabilities, its physical size is much greater. The increase in size includes not only the actual tube elements but also the glass insulation and the leads. As a consequence, for frequencies in the neighborhood of 50 megacycles, this tube does not fulfill the condition that the length of the tube be a small fraction of the wavelength. Fig. 16.27 shows the approximate equivalent network representing the tube. In order to permit neutralization, irrespective of the inductance of the grid lead, the tube is so constructed that a neutralizing sleeve, described in the preceding section, can be placed over the insulating wall of the tube and give almost direct capacity coupling to the grid element itself. It should also be apparent that, because of the great length of the filament lead, the cathode cannot usually be grounded directly, but rather must be grounded through a half-wave line.

In spite of these size limitations, these tubes have given fairly good service as the output stages in a television transmitter.

Included with tube type RCA 899 in Fig. 16.26 are two power tubes lying between the RCA 888 and the tube just described. These are the RCA 858 rated at 20 kilowatts, and the RCA 846 capable of dissipating 2.5 kilowatts.

Although not yet available for practical television transmitter work, new tubes better suited for ultra-high-frequency broad-band power amplification are being developed in research laboratories all over the world. The final form of these tubes cannot, of course, be given, but certain trends are in evidence.

The application of electron optics to power tube design will permit

beaming the electrons to such an extent that screen and suppressor grids become possible. This will reduce the effective capacity, in both the input and output sides of the tube, by eliminating or reducing the size of the neutralizing condensers.

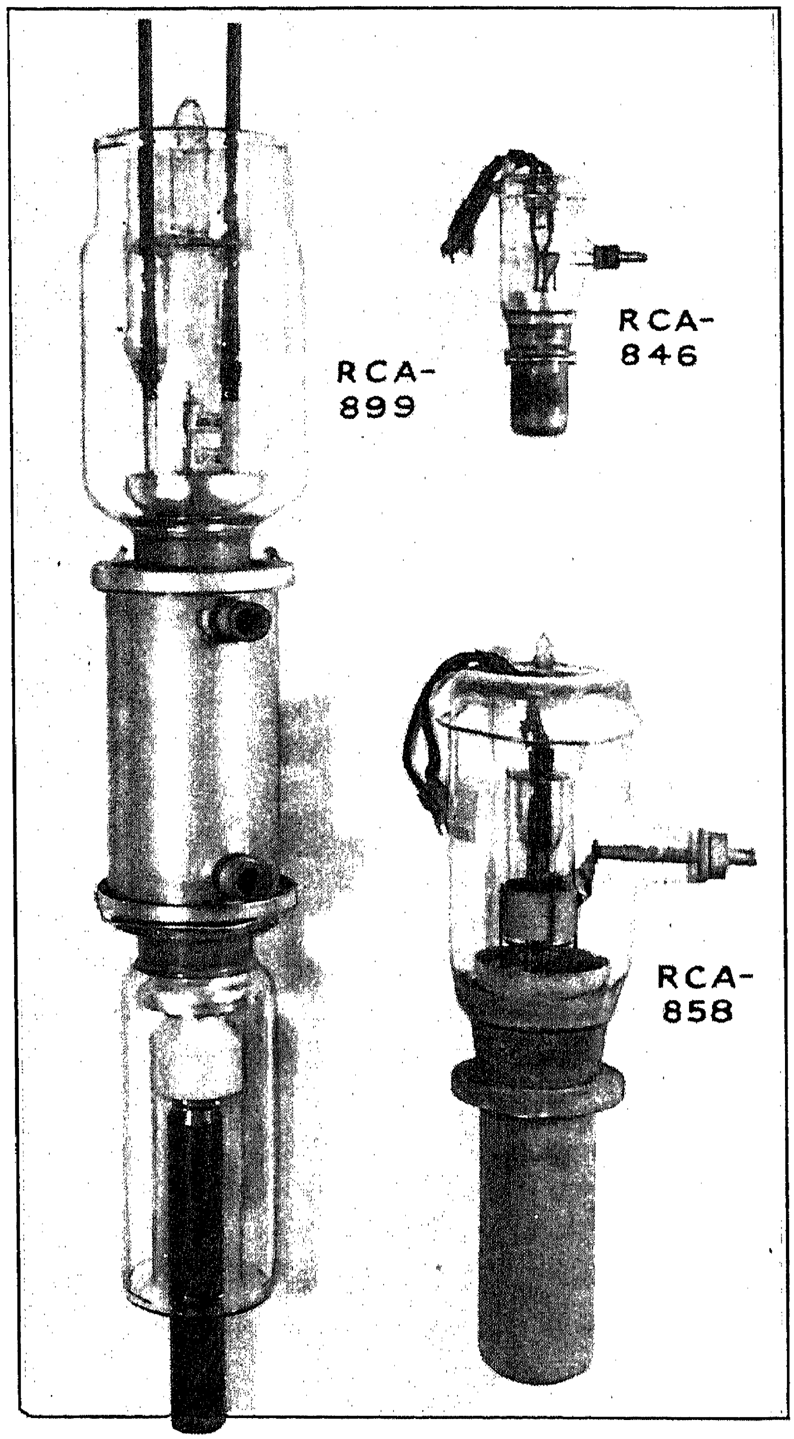


Fig. 16.26.—Ultra-High-Frequency Transmitting Tubes.

Cathodes improved both as to geometry and emitting materials will permit obtaining greater emission for a given cathode size, and will also allow the cathode to be operated at a lower temperature. This, in turn,

simplifies the problem of cooling the grid, making possible, for example, small radiation-cooled tungsten control grids.

Better contact between the cooling fluid and the anode will aid in reducing the physical size of the anode.

Finally, the use of demountable power tubes may greatly simplify the problem of making the leads to the various elements of minimum length.

These improvements will have the effect of increasing the efficiency of the present types of modulating and power-amplifying systems by a factor of two or three.

It is not at all improbable that in the not too distant future entirely new forms of tubes may come into being for ultra-high-frequency work. Such a possibility is clearly indicated in the present very interesting

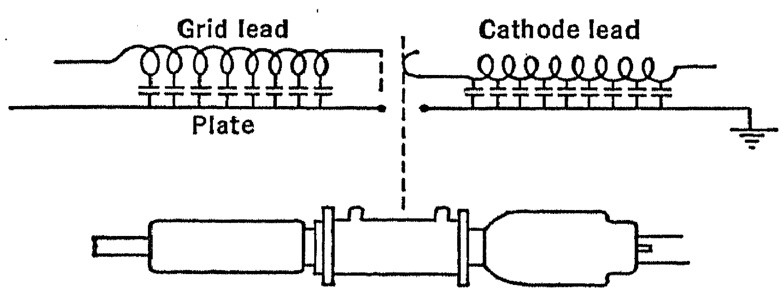


FIG. 16.27.—Approximate Equivalent Network of Electrodes of High-Power Ultra-High-Frequency Transmitting Tube.

research work that is being done on tubes whose output is the product, not of the electron current itself, but of the displacement current resulting from the electron motion. These tubes are exemplified by the Barkhausen oscillator, the Heil* oscillator, the velocity-modulated amplifiers reported on by Hahn and Metcalf,† and the inductive output tube of Haeff.‡

16.6. The Transmission Line. The modulated ultra-high-frequency output from the final power stages of the transmitter is radiated from an antenna which, in general, is located at some distance from the generating units. Therefore the problem arises of carrying the signal over a long length of transmission line without phase or amplitude distortion and without too great loss of power. This problem differs from that of the transfer of power at low frequencies in that the wavelength of the electrical oscillation is short compared with the length of the line under consideration. The voltage at any instant between the pair of conductors making up the line varies from point to point, going through a series of maxima and minima, and the instantaneous current likewise undergoes

* See Heil, reference 14.

† See Hahn and Metcalf, reference 15.

‡ See Haeff, reference 16.

a series of variations. A typical transmission line is represented schematically in Fig. 16.28a. Between such a pair of conductors there is a certain capacity C_0 per unit length of the line, and the line itself has inductance per unit length to be designated by L_0. Furthermore, since the metal wires are not perfect conductors, nor the insulation separating the wires a perfect insulator, there will be a resistance R_0 per unit length, and a conductance G_0 between them. The line may be represented by the equivalent circuit given in Fig. 16.28b.

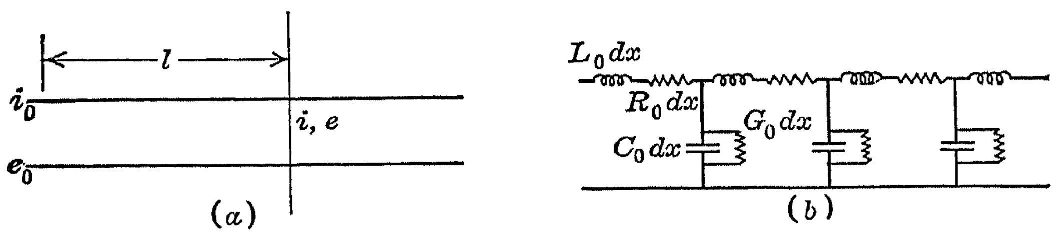


FIG. 16.28.—Transmission Line and Its Effective Circuit Equivalent.

Considering an elemental length of the line, the current and voltage will change by increments

$$de = -\ i(R_0 + j\omega L_0)dx, \tag{16.9a}$$
$$di = -\ e(G_0 + j\omega C_0)dx. \tag{16.9b}$$

A complete solution of these equations describes the behavior of the line. However, such a solution is unnecessarily complicated for the present discussion. In any practical line used for short-wave power transmission the leakage conductance can be neglected, and furthermore $R_0/\omega L_0$ is very small compared with unity. Under these conditions the voltage e and current i at a distance l from the input end of a semi-infinite line are given by the following relations:

$$e = e_1\epsilon^{-\alpha l}\epsilon^{-j\beta l}, \tag{16.10a}$$
$$i = i_1\epsilon^{-\alpha l}\epsilon^{-j\beta l}, \tag{16.10b}$$

where e_1 and i_1 are the input voltage and current and

$$\alpha = \frac{R_0}{2}\sqrt{\frac{C_0}{L_0}},$$

$$\beta = \omega L_0\sqrt{\frac{C_0}{L_0}}.$$

It might be noted here that, since the velocity of propagation along the line is:

$$v = \frac{1}{\sqrt{L_0 C_0}}, \tag{16.11}$$

and the wavelength along the line

$$\lambda = \frac{2\pi v}{\omega},$$

the phase factor can be expressed as

$$\beta = \frac{2\pi}{\lambda}.$$

The input impedance of the semi-infinite line can, with the aid of the above equations, be shown to be:

$$Z_s = \sqrt{\frac{L_0}{C_0}}\left(1 - j\frac{R_0}{2\omega L_0}\right). \tag{16.12}$$

This impedance is known as the characteristic or surge impedance* of the line, and is, of course, the impedance of a finite line terminated in an impedance equal to the surge impedance. For any practical line the imaginary term $j(R_0/2\omega L_0)$ will be less than 0.1 per cent. It can therefore be omitted in most considerations. Under certain resonant conditions which are discussed later it plays an important role, however, in determining the properties of the line.

When a finite line is terminated in any impedance Z_2 other than its surge impedance Z_s, it may be demonstrated to have the following characteristics:

$$Z_1 = Z_s\frac{Z_2 \cosh ml + Z_s \sinh ml}{Z_2 \sinh ml + Z_s \cosh ml}, \tag{16.13a}$$

$$i_2 = i_1\frac{Z_s}{Z_2 \sinh ml + Z_s \cosh ml}, \tag{16.13b}$$

$$e_2 = e_1\frac{Z_s}{Z_2 \cosh ml + Z_s \sinh ml}, \tag{16.13c}$$

* The characteristic impedance when the leakage conductance is considered is

$$Z_s = \sqrt{\frac{R_o + j\omega L_o}{G_o + j\omega C_o}}.$$

Some authors distinguish between surge impedance and characteristic impedance, defining the terms as $\sqrt{\frac{L_o}{C_o}}$ and $\sqrt{\frac{R_o + j\omega L_o}{G_o + j\omega C_o}}$, respectively. This distinction is not made in the present discussion, since $\sqrt{L_o/C_o}$ can always be considered as an approximation of $\sqrt{\frac{R_o + j\omega L_o}{G_o + j\omega C_o}}$.

where the subscripts 1 and 2 refer to the input and output ends of the line, respectively, and

$$m = \left(\frac{R_0}{2\omega L_0} + j\right)\frac{2\pi}{\lambda}.$$

The characteristics listed in Eqs. 16.13*a* to 16.13*c* are due to the fact that the abrupt change in impedance at the termination produces reflections back along the line. When the line length is a simple fraction of the wavelength it has certain interesting and valuable properties. From a practical standpoint the most important of these special lines are those having lengths given by $l = \lambda/2$, $\lambda/4$, and $\lambda/8$.

A half-wavelength line, when it is short-circuited, has an input impedance which can be expressed as:

$$Z_1 = \frac{R_0\lambda}{4}, \tag{16.14}$$

and has properties of a series resonant circuit. When a similar line is terminated with an infinite impedance, the input impedance becomes:

$$Z_1 = \frac{4L_0}{R_0\lambda C_0}, \tag{16.15}$$

and the circuit performs like an anti-resonant circuit. Neglecting the small quadrature component of the surge impedance and writing $Z_s \doteq \sqrt{L_0/C_0}$ a half-wave line behaves in such a way that the terminal impedance is transferred to the input impedance, i.e.:

$$Z_1 \cong Z_2, \tag{16.16}$$

the approximation requiring that Z_2 be not too far different from Z_s.

A quarter-wavelength line has the following properties: The input impedance is

$$Z_1 = \frac{8L_0}{R_0\lambda C_0}, \text{ when } Z_2 = 0 \tag{16.17}$$

and the circuit is anti-resonant. On the other hand, when $Z_2 = \infty$, the input impedance can be written as:

$$Z_1 = \frac{R_0\lambda}{8}, \tag{16.18}$$

and the circuit is resonant. Under the same assumptions which led to impedance transfer for a half-wave line, the quarter-wave line becomes an impedance inverter, i.e.:

$$Z_1 = \frac{Z_s^2}{Z_2}. \tag{16.19}$$

An open-circuited line whose length is equal to $\lambda/8$ has an impedance equal to the surge impedance, and a leading phase angle of 90°. A short-circuited line of similar length has the same impedance, but a lagging phase angle. Similarly very short lines under these conditions become more reactive in their behavior, approaching as a limit pure capacities and inductances having a magnitude $C_0 l$ and $L_0 l$, respectively.

These special lines are very important in the design of ultra-high-frequency tank circuits, transformers, and filters. Examples of the use of these transmission line equivalents of ordinary circuit elements will be found in the concentric line tank circuit described above and in the filter

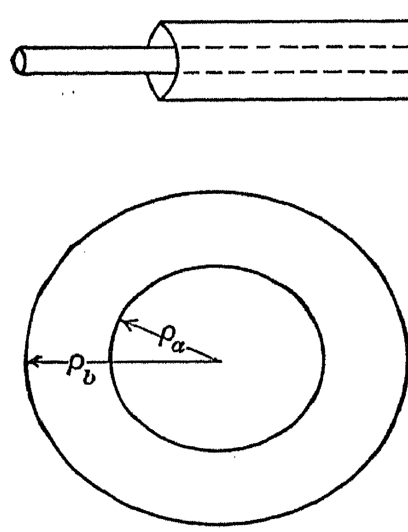


Fig. 16.29.—Concentric Line.

for removing the unwanted sideband in single sideband transmission, discussed later in this section.

The coaxial or concentric line is a very important special case of the transmission line. It consists of two concentric cylindrical conductors, insulated from each other with low-loss spacers, so that for all practical purposes the dielectric medium between them is air. Because of its high efficiency due to low radiation loss, this type of line is used almost exclusively in ultra-high-frequency power transfer. A coaxial line is illustrated in cross-section in Fig. 16.29. The surge impedance of such a line can be calculated directly, if the assumption is made that skin effect causes most of the current to flow on the outside of the conductors, so that internal inductance plays no role (a condition almost exactly fulfilled in an ultra-high-frequency cable). Use is made of the fact that the velocity of propagation along the cable is equal to the velocity of light, and of Eq. 16.11

$$\sqrt{L_0 C_0} = \frac{1}{c}, \text{ where } c = 3 \times 10^{10} \text{ cm/sec.}$$

The surge impedance is therefore given by:

$$Z_s = \sqrt{\frac{L_0}{C_0}} = \frac{1}{cC_o} = \frac{1}{c}\int_{\rho_a}^{\rho_b}\frac{2dr}{r} \tag{16.20}$$

$$= \frac{2}{3 \times 10^{-10}} \ln \frac{\rho_b}{\rho_a} \text{ (esu)}$$

$$\text{or } 60 \ln \frac{\rho_b}{\rho_a} \text{ (ohms)},$$

ρ_a and ρ_b being the radii of the inner and outer conductor respectively.

The resistance per unit length R_0 determines the power loss and also many of the electrical characteristics of the cable or transmission line. This resistance is not, however, simply the product of the d-c resistivity and the cross-sectional area of the conductor because, as has already been mentioned, the ultra-high-frequency current is all concentrated within a small fraction of an inch of the surface. A fairly accurate value can be obtained for the "skin-effect" resistance of a cylindrical conductor at ultra-high frequency from the formula

$$R_0 = \frac{3.16 \times 10^{-5}}{\rho}\sqrt{\sigma f} \text{ ohms/cm}, \tag{16.21}$$

where ρ is the radius (cm) and σ the specific resistivity (ohm-cm). In addition to the resistance of the conductor itself, radiation and dielectric loss may add to R_0 for certain types of lines, but these contribute a negligible amount in a well-designed coaxial line. It is interesting to note that the resistance varies inversely with the radius, rather than with the inverse square of the radius, as it does at low frequencies.

The optimum radius for the inner conductor, based on the most efficient power transfer, can be readily determined for a concentric line of given radius of outer conductor. Let I be the current flowing; then the power loss is given by

$$P_L = I^2(R_b + R_a),$$

where $R_b = k/\rho_b$ is the resistance of the outer conductor and $R_a = k/\rho_a$, that of the inner. The power transfer is

$$P = I^2 Z_s.$$

The efficiency η can therefore be expressed as:

$$\eta = \frac{P}{P_L} = \frac{Z_s}{R_b + R_a} = \frac{60 \ln \rho_b/\rho_a}{k\left(\frac{1}{\rho_b} + \frac{1}{\rho_a}\right)}. \tag{16.22}$$

Differentiating with respect to ρ_b/ρ_a and equating to zero gives:

$$\ln \frac{\rho_b}{\rho_a} = 1 + \frac{\rho_a}{\rho_b},$$

which leads to a value

$$\frac{\rho_b}{\rho_a} \cong 3.6$$

irrespective of the other constants of the line. This, of course, is the same radius ratio mentioned in connection with concentric tank circuits in an earlier section.

The surge impedance of a line having a 3.6 radius ratio is, from Eq. 16.20, found to be 76.6 ohms. Other things being equal, it is, of course, advisable to use a line having approximately this impedance. However, as can be seen from Eq. 16.22, the efficiency curve has a very broad maximum. Lines having surge impedances ranging from 50 to 125 ohms are almost equally effective as a means of power transfer. In practice usually some other consideration will determine what value of surge impedance within these limits is most suited to the application in question. Problems of impedance matching are, in general, the determining factor.

Concentric lines of the type described are used in practically all television installations to convey the power from the transmitter to the antenna. When the line is terminated with its own surge impedance, the input impedance will, from Eq. 16.20, be independent of the frequency of the signal. The attenuation of the output voltage is not entirely frequency independent as is evidenced by the fact that the resistance from Eq. 16.21, appearing in Eq. 16.13*c*, is a function of frequency. However, since this resistance depends upon the square root of the frequency the variation over the bandwidth used in practice is relatively unimportant. The phase of the sideband frequency relative to the carrier is not fixed but depends upon the length of the line and the frequency. However, the variation is linear with frequency and therefore does not introduce distortion in the video signal. These characteristics, together with its high efficiency, make a correctly terminated coaxial cable well suited to the transfer of television power.

When the transmission line terminates in an antenna, the impedance of the antenna must match the surge impedance of the line. Unless this match is fairly exact, reflections at the termination will introduce serious phase and amplitude distortion. It does not suffice to terminate the line in a load which equals the surge impedance at the carrier frequency alone, but rather the termination must match the line at all frequencies

transmitted. The problem of matching the cable involves the load characteristic of the antenna and will be considered briefly in the next section.

As has already been mentioned, circuit elements are often inserted in a transmission line to filter out unwanted components of the frequency

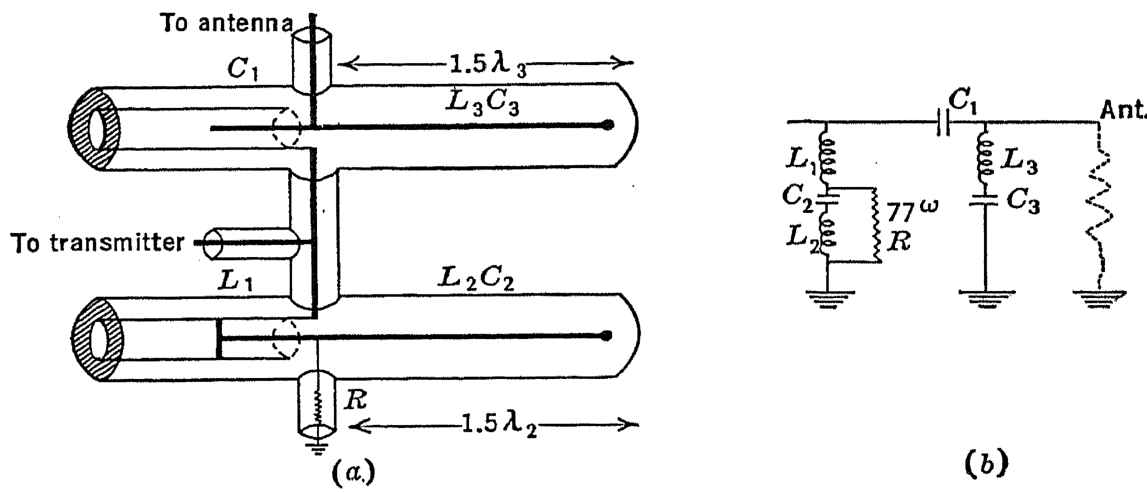


FIG. 16.30.—Coaxial Line High-Pass Filter.

spectrum. A filter is required for single-sideband transmission where both sidebands are produced at the modulator. The filter for this purpose must be dissipative and must match the surge impedance of both the cable from the transmitter and that from the antenna if reflections are to be avoided. Fig. 16.30 illustrates a filter element for sideband removal designed by G. H. Brown. This filter is in reality a high-pass filter between the antenna and the transmitter, whose circuit equivalent is shown in Fig. 16.30*b*. The circuit constants are so chosen that L_2 and C_2 resonate at the frequency f_0, which is equal to or slightly higher than carrier frequency. L_3 and C_3 resonate at f_c, about 2 megacycles lower than f_0. At f_0 the network becomes in effect a π section terminated by the antenna, while for the lower frequency f_c it is a π section supplying the dissipating resistance R. The circuit presents an impedance of approximately 77 ohms to the line over the entire frequency range.

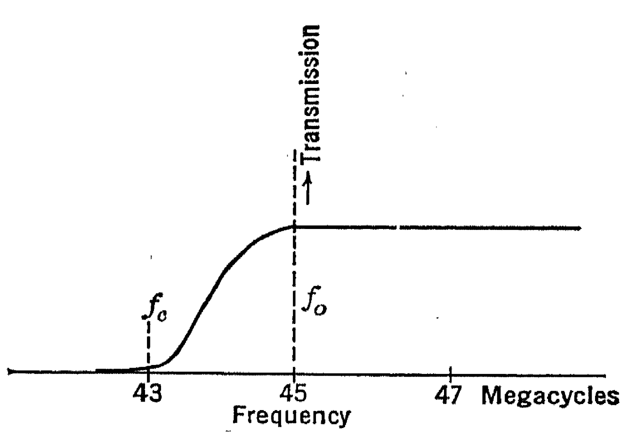


FIG. 16.31.—Characteristic of Filter Shown in Fig. 16.30.

In the actual television transmitter such a filter is made up of concentric lines having characteristics corresponding to the elements in the

equivalent circuit. The coaxial line filter is shown in Fig. 16.30*a*, where the resonant circuits are replaced by half-wave (or three half-wavelength) lines, and the capacitance and inductance by open and closed one-eighth-wave lines. The transmission characteristic of this filter is shown in Fig. 16.31. Where a sharper cutoff is required several of these filters may be used in series, and sharp band reject filters may be added to them.

16.7. The Antenna. The antenna is the means of coupling the output of the transmitter to the medium through which electromagnetic waves are propagated. Its operation follows directly from the Maxwell equations for the electromagnetic field, which predict that an accelerated electric charge will give rise to radiation which is propagated at the velocity of light. In the case of an antenna, the accelerated charges are the oscillatory current flowing in the antenna conductors.

It can be shown that an element of wire of length dl, carrying a current I, whose frequency is f, will produce a field strength E given by:

$$dE = \frac{60\pi}{xc} fIdl \cos 2\pi f\left(t - \frac{x}{c}\right) \cos \theta. \tag{16.23}$$

In this equation x is the distance from the element in meters and θ the angle between the direction of x and the plane normal to dl, E being measured in volts per meter. If the conductor has finite length, the radiation at any point can be determined by adding up the radiation (taking into account magnitude, phase, and direction of the electric vector) from every element of the conductor. It is obvious that the way in which the radiation is distributed about the conductor, or in other words the radiation pattern, depends upon its geometry and current distribution. Determining the radiation pattern is really an interference or diffraction problem.

A simple but very important antenna for ultra-high frequency is that formed by a straight wire located a number of wavelengths above the ground whose length is equal to a half-wavelength. Such an antenna is known as a dipole. The current and voltage in such a conductor form standing waves, giving the distribution shown in Fig. 16.32. The radiation pattern is included in the figure and will be seen to follow fairly closely the cosine law indicated by Eq. 16.23, as is to be expected from the fact that the current is concentrated chiefly in a small length near the center of the dipole. The electrical properties of a half-wave dipole are essentially those of a resonant circuit. If the radiator is divided in the middle at the current maximum and the two halves connected to the ends of a transmission line, the termination of the line can be represented as a series circuit consisting of an inductance, a condenser, and a re-

sistor. The value of the resistance is determined by the power which is radiated from the antenna for a given current from the line, and is known as the radiation resistance. It is equal to a resistance which will dissipate the radiated power when a current equal to the antenna current flows through it.

The radiation resistance of a half-wave dipole is approximately 73.2 ohms. As the length is increased, the impedance increases, having a large inductive component which goes through a maximum and then decreases until at one wavelength the antenna is essentially a parallel resonant circuit presenting a large purely resistive impedance. At three-halves wavelengths the antenna is again series resonant with a radiation re-

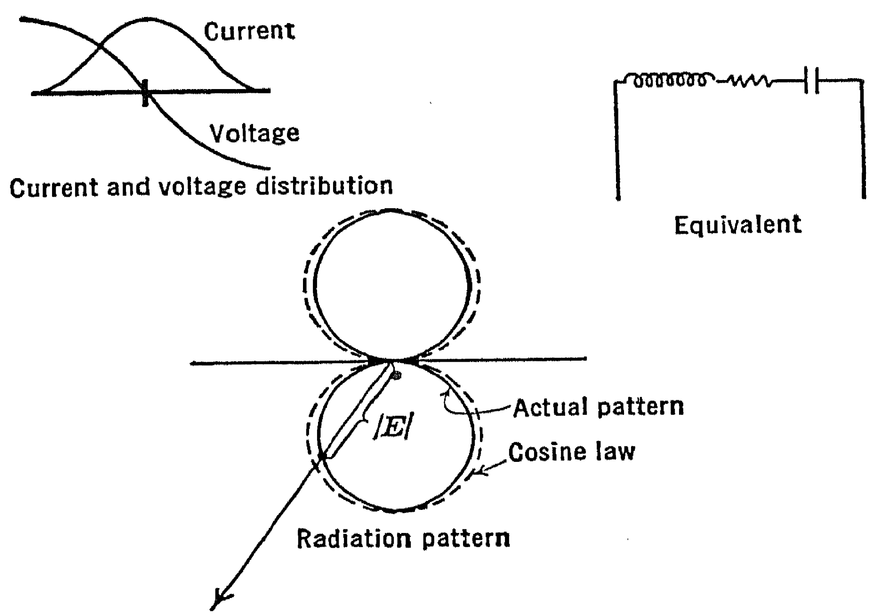


Fig. 16.32.—Properties of a Half-Wave Dipole Antenna.

sistance in the neighborhood of 150 ohms. Decreasing the length of an antenna from a half-wavelength increases its impedance by adding a capacitative component. Obviously a variation in frequency produces an effect on the impedance similar to lengthening or shortening the antenna. This variation in magnitude and phase of the impedance of an antenna with frequency introduces a difficulty when the antenna terminates a transmission line in that it matches the surge impedance of the line at only one frequency. To overcome this for wide band transmission, such as involved in television broadcasting, it is necessary to employ either special matching circuits or an antenna which has a flatter impedance characteristic. In actual practice both methods are used.

The half-wave dipole is usually excited by two concentric lines, each feeding half of the antenna as shown in Fig. 16.33. The terminating impedance of each line will only be 36 ohms, so that an impedance trans-

former must be used to match a line of higher impedance accurately, for example, a 70-ohm cable. A number of impedance matching arrangements are possible. One is shown in Fig. 16.34, involving quarter-wave impedance inverters described in the preceding section. While the terminating impedance at carrier frequency can be matched to the 70-ohm cable with a single inverter, the double arrangement shown gives a much better match at sideband frequencies.

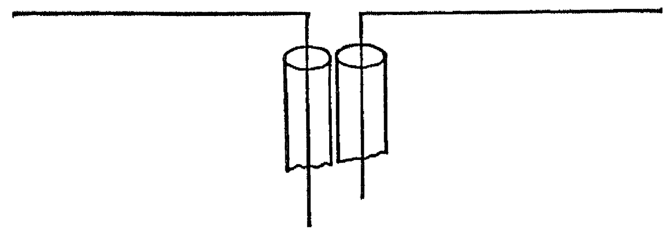


Fig. 16.33.—Method of Feeding Antenna.

A single cable often carries the power from the transmitter. This cable must be divided to form the feeders to the two separate sides of the antenna. The division must be made in such a way that it does not introduce a mismatch which will produce reflections. Two dividers meeting this requirement are shown in Fig. 16.35.

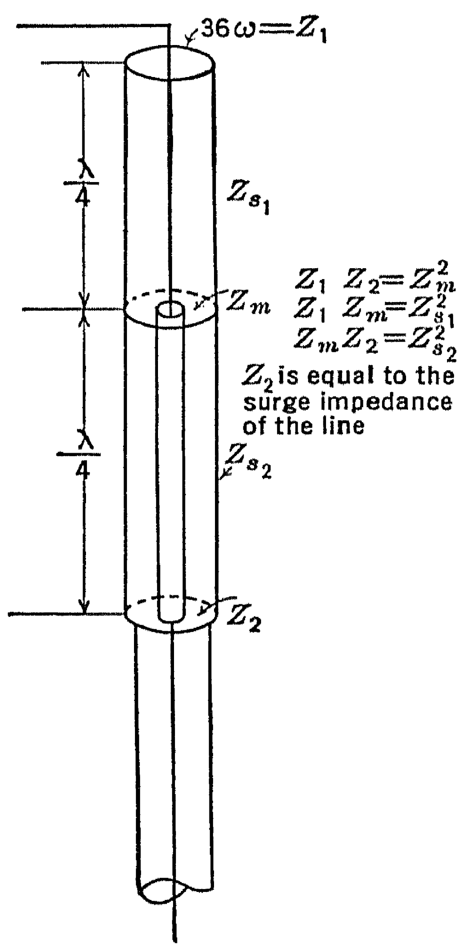


Fig. 16.34.—Coaxial Matching Transformer.

There are many types of ultra-high-frequency antennas, chosen to give the desired radiation pattern and load characteristics. A large class is made up of arrays of linear

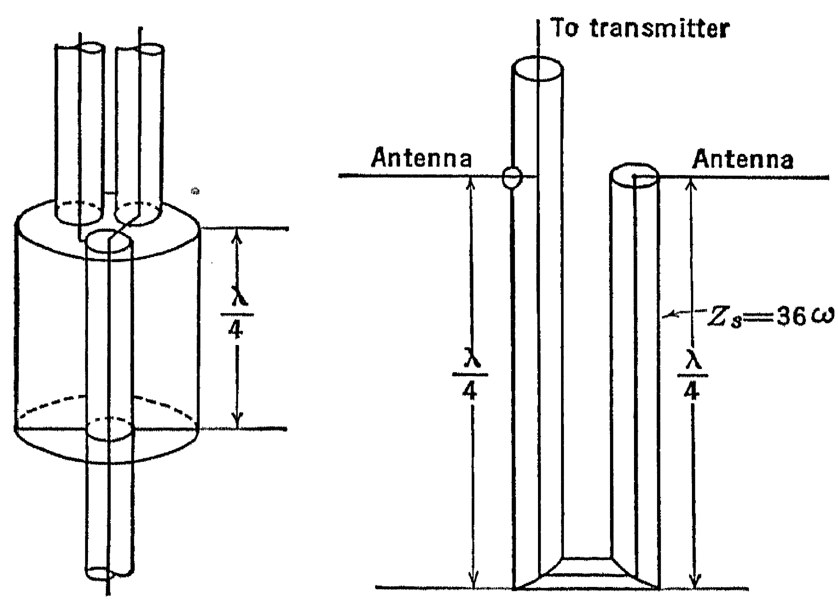


Fig. 16.35.—Coaxial Transformer for Line Division.

dipoles suitably spaced and oriented, fed so that currents in the various elements have a definite phase relation to one another. The radiation from all the elements is coherent and an interference pattern is formed

which can be calculated by methods similar to those for calculating the interference of sound, light, or X-rays. Two antennas of this class are shown in Fig. 16.36, one being a triangular array of the type used for some of the television transmission tests made from the Empire State Building, while the second is a so-called turnstile antenna.* Both antennas are made up of stacks of parallel horizontal dipoles, excited so that the currents in the elements constituting each stack are equal and in phase. This directs the radiation strongly into a horizontal plane, and consequently the power emitted upward, where it would be wasted, is very small. The current fed to the two stacks is in quadrature in the turnstile antenna. In the triangular array the current supplied to the three stacks is in phase. Both these antennas produce a fairly uniform pattern confined to a horizontal plane. Like the single horizontal dipole, both these antennas are strongly resonant and therefore require a special matching element to couple them to a coaxial line.

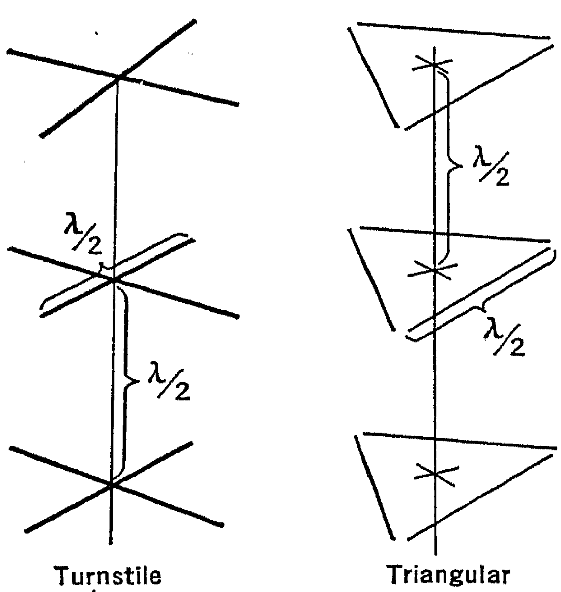


FIG. 16.36.—Turnstile and Triangular Antennas.

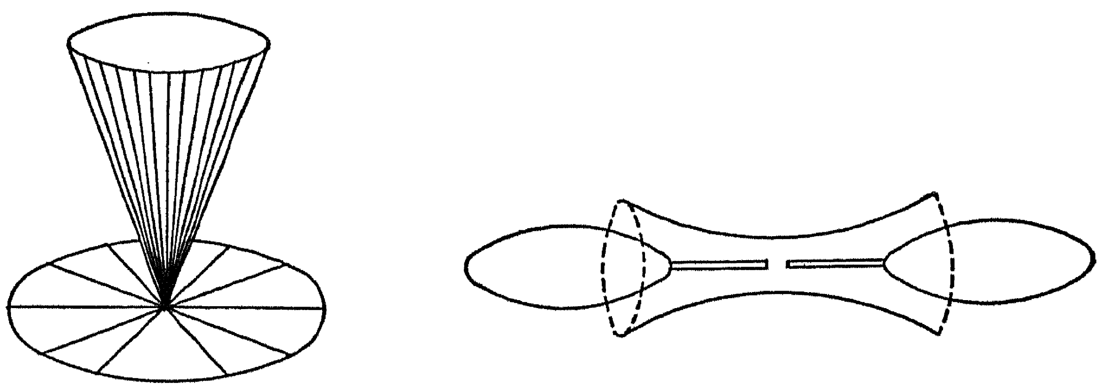
A second class of antennas is constituted of those whose radiating element or elements are not simple linear conductors. The calculation of the performance of this type of antenna is much more difficult because a simple sinusoidal current distribution cannot be assumed. Therefore in addition to the interference problem there arises that of determining the current flow. Two radiating elements of this type are shown in Fig. 16.37*a* and *b*. Both these have impedance characteristics which are much less dependent upon frequency than are those of antennas made up of linear dipoles. The antenna illustrated in Fig. 16.37*a* produces a vertically polarized wave which is directed uniformly in a horizontal plane. Horizontally polarized radiation is obtained from the antenna shown in Fig. 16.37*b*.† In order to obtain a uniform pattern with this radiator the antenna must consist of two such elements at right angles.

It will be readily apparent from the foregoing that the field of the ultra-high-frequency antenna and associated coaxial coupling system is one of extreme complexity, and that in a brief discussion such as given above only the nature of the problem can be outlined.

* See G. H. Brown, reference 20.

† See N. E. Lindenblad, reference 21.

16.8. Propagation at Ultra-High Frequencies. Energy leaves the antenna in the form of electromagnetic radiation. This radiation is polarized with its electric vector lying in a plane including the direction of current flow in the antenna. Therefore the electromagnetic waves propagated from a horizontal dipole are horizontally polarized, while those from a vertical dipole have their electric vector lying in a vertical plane. Both vertically and horizontally polarized ultra-high-frequency signals are suitable for communication purposes, and each is employed as a means of transmitting television signals.



Vertical conical antenna (*a*) Horizontal radiator (*b*)

Fig. 16.37.—Special Wide-Band Television Antennas.

The field strength E at a distance d from a half-wave dipole in a plane about its center is given by

$$E = \frac{7\sqrt{W}}{d}, \tag{16.24}$$

where W is the power radiated in watts and d and E are in meters and volts per meter, respectively. This relation is true only for a radiator which is in free space, in other words when no reflected radiation is present. The field strength in other directions varies nearly with the cosine of the angle between a plane normal to the dipole and the line joining the center of the antenna with the point at which the electric vector is measured. This has already been mentioned in connection with the radiation pattern from such a radiator in the preceding section.

When the antenna and the point of measurement are close to the ground, this relation is no longer true because of the interference effect between the direct radiation and that reflected by the earth's surface. An approximate expression for the effect of this interference can be easily found. Fig. 16.38 represents a transmitting and receiving antenna separated by a distance d. The heights of these antennas are h and a, respec-

tively, these heights being small compared with the distance d. The instantaneous field strength at a due to direct radiation is:

$$e_1 = \frac{7\sqrt{W}}{x_1} \cos 2\pi\left(ft + \frac{x_1}{\lambda}\right), \tag{16.25}$$

where x_1 is the path distance. The reflected radiation reaches the antenna over a slightly longer path x_2, and furthermore suffers a phase reversal of 180° upon reflection at the earth's surface. Therefore the field strength due to it is:

$$e_2 = -\frac{7\sqrt{W}}{x_2} \cos 2\pi\left(ft + \frac{x_2}{\lambda}\right). \tag{16.25a}$$

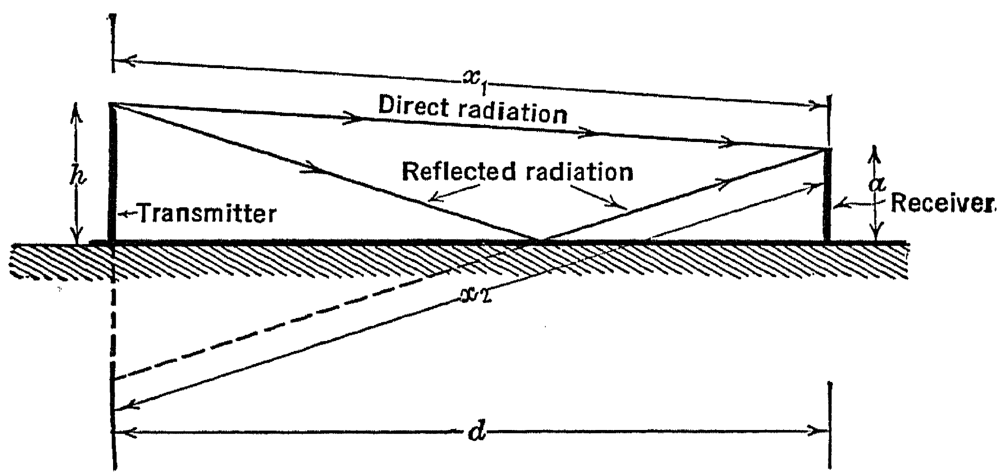


FIG. 16.38.—Determination of Signal Intensity at Receiver.

The resultant field, which is the vector sum of e_1 and e_2, can be written as:

$$\begin{aligned} e &= \frac{7\sqrt{W}}{x_1} \cos 2\pi\left(ft + \frac{x_1}{\lambda}\right) - \frac{7\sqrt{W}}{x_2} \cos 2\pi\left(ft + \frac{x_2}{\lambda}\right) \\ &= \frac{7\sqrt{W}}{d}\, 2 \sin 2\pi\left(ft + \frac{d}{\lambda}\right) \sin \pi\left(\frac{x_2 - x_1}{\lambda}\right) \\ &\cong \frac{7\sqrt{W}}{d}\, 2\pi \frac{x_2 - x_1}{\lambda} \sin 2\pi\left(ft + \frac{d}{\lambda}\right). \end{aligned} \tag{16.26}$$

It can be seen from the figure that x_1 and x_2 are given approximately by:

$$x_1 = d + \frac{1}{2}\frac{(h - a)^2}{d},$$

$$x_2 = d + \frac{1}{2}\frac{(h + a)^2}{d},$$

and

$$x_2 - x_1 = \frac{2ah}{d}.$$

Therefore the magnitude of the field strength at the receiving antenna is:

$$E = \frac{28\pi\sqrt{Wah}}{d^2\lambda}, \tag{16.27}$$

or, if a and h are in feet, d in miles, E in millivolts per meter, and λ in meters:

$$E = \frac{0.0032ah\sqrt{W}}{d^2\lambda}. \tag{16.27a}$$

It is interesting to note that the field strength varies inversely with the square of the distance separating the antennas, and furthermore that it increases with increasing frequency. This relation is found to hold very accurately over distances which are within the horizon and over terrain which is free from obstructions. In urban districts absorption and reflection of the radiation are found to cause marked departures from the values given by the formula. In general the field strengths for radiation in the neighborhood of 50 megacycles will be found to be 30 to 60 per cent lower than the values given by Eq. 16.27a, and at higher frequencies the departure will be even greater. Of course, there will be regions of reinforcing reflections where the field strength is greater and others where it is lower.

Beyond the horizon the field strength decreases more rapidly than the inverse square, owing to the quasi-optical properties of the radiation. At 40 megacycles the field strength decreases with $1/D^{3.6}$; at 100 megacycles it is given by approximately the inverse fifth power; and at 400 megacycles by the inverse ninth power. The only radiation reaching points beyond the line of sight is thought to be that diffracted by the earth's surface at the horizon, as the calculated results obtained on this basis agree fairly well with experimental measurements.

The effective horizon, determined by the height of the transmitting antenna, is given by:

$$l = 1.22\sqrt{h},$$

where h is the antenna height in feet and l is the distance to the horizon in miles. Usually the value obtained from this expression is taken as the practical service radius of a given transmitter, although reliable reception can be had at somewhat greater distance.

Very little ultra-high frequency is reflected from the ionic layers of the upper atmosphere, so that reception at a great distance is a rare occurrence. However, it is not impossible, and occasionally a television picture transmitted from London is received in the United States, although usually badly distorted as a result of multiple paths.

REFERENCES

1. F. E. Terman, "Radio Engineering," McGraw-Hill, New York, 1937.
2. F. Schröter (editor), "Fernsehen," J. Springer, Berlin, 1937.
3. H. Pender and K. McIlwain, "Electrical Engineers' Handbook—Electric Communication and Electronics," John Wiley & Sons, New York, 1936.
4. E. W. Engstrom and C. M. Burrill, "Frequency Assignments for Television," *R. C. A. Rev.*, Vol. 1, pp. 88–93, January, 1937.
5. R. D. Kell, A. V. Bedford, and M. A. Trainer, "An Experimental Television System, Part II, Transmitter," *Proc. I. R. E.*, Vol. 22, pp. 1246–1265, November, 1934.
6. F. E. Terman and W. C. Roake, "Calculation and Design of Class C Amplifiers," *Proc. I. R. E.*, Vol. 24, pp. 620–632, April, 1936.
7. J. W. Conklin and H. E. Gihring, "Television Transmitter Operating at High Powers and Ultra High Frequencies," *R. C. A. Rev.*, Vol. 2, pp. 30–44, July, 1937.
8. C. W. Hansel and P. S. Carter, "Frequency Control by Low Power Factor Line Circuits," *Proc. I. R. E.*, Vol. 24, pp. 597–619, April, 1936.
9. N. E. Davis and E. Green, "The Marconi EMI Television System, Part III, The Radio Transmitter," *Jour. I. E. E.*, Vol. 83, pp. 782–792, December, 1938.
10. W. N. Parker, "A Unique Method of Modulation for High-Fidelity Television Transmitters," *Proc. I. R. E.*, Vol. 26, pp. 946–962, August, 1938.
11. P. J. H. A. Nordlohne, "Experimental Short Wave Broadcasting Station PCJ," *Philips Tech. Rev.*, Vol. 3, pp. 17–27, January, 1938.
12. W. G. Wagener, "The Developmental Problem and Operating Characteristics of Two New Ultra-High Frequency Triodes," *Proc. I. R. E.*, Vol. 26, pp. 401–414, April, 1938.
13. W. G. Wagener, "The Requirements and Performance of a New Ultra High Frequency Power Tube," *R. C. A. Rev.*, Vol. 1, pp. 258–264, October, 1937.
14. A. A. Heil and O. Heil, "Production of High Frequency Electric Waves of High Intensity," *Z. Physik*, Vol. 95, pp. 752–762, 1935.
15. W. C. Hahn and G. F. Metcalf, "Velocity Modulated Tubes," *Proc. I. R. E.*, Vol. 27, pp. 106–116, February, 1939.
16. A. V. Haeff, "Ultra High Frequency Power Amplifier of Novel Design," *Electronics*, Vol. 12, pp. 30–32, February, 1939.
17. E. J. Stirba and C. B. Feldman, "Transmission Lines for Short Wave Radio Systems," *Proc. I. R. E.*, Vol. 20, pp. 1163–1202, July, 1932.
18. L. S. Nergaard, "Survey of Ultra-High Frequency Measurements," *R. C. A. Rev.*, Vol. 3, pp. 156–195, October, 1938.
19. Lester F. Reukima, "Transmission Lines at Very High Frequencies," *Elect. Eng.*, Vol. 56, pp. 1002–1011, August, 1937.
20. G. H. Brown, "The Turnstile Antenna," *Electronics*, Vol. 9, pp. 14–17 and 48, April, 1936.
21. N. E. Lindenblad, "Television Transmitting Antenna for Empire State Building," *R. C. A. Rev.*, Vol. 3, pp. 387–408, April, 1939.
22. H. H. Beverage, "Some Notes on Ultra High Frequency Propagation," *R. C. A. Rev.*, Vol. 1, pp. 76–87, January, 1937.
23. B. Trevor and P. S. Carter, "Notes on Propagation of Waves Below Ten Meters in Length," *Proc. I. R. E.*, Vol. 21, pp. 387–426, March, 1933.

24. C. R. BURROWS, L. E. HUNT, and A. DECINO, "Ultra-Short-Wave Propagation Mobile Urban Transmission Characteristics," *Bell Sys. Tech. J.*, Vol. 14, pp. 253–272, April, 1935. See also, J. C. SCHELLING, C. R. BURROWS, E. B. FERRELL, *Bell Sys. Tech. J.*, Vol. 12, pp. 125–161, April, 1933.

25. R. W. GEORGE, "A Study of Ultra-High Frequency Wide Band Propagation Characteristics," *Proc. I. R. E.*, Vol. 27, pp. 28–35, January, 1939.

26. R. S. HOLMES and A. H. TURNER, "An Urban Field Strength Survey at Thirty to One Hundred Megacycles," *Proc. I. R. E.*, Vol. 24, pp. 755–770, May, 1936.

27. D. R. GODDARD, "Observations on Sky Wave Transmission on Frequencies above 40 Megacycles," *R. C. A. Rev.*, Vol. 3, pp. 309–315, January, 1939.

CHAPTER 17

THE RECEIVER

The present chapter discusses the more important aspects of the receiver design, considering the circuits involved and the part they play in the reproduction of the picture. In part it is merely an integration of the discussion of the preceding chapters, tying together such elements as the Kinescope, the video amplifier, and the deflection generator into a coordinated whole. There are also a number of units in the receiver which are as yet undiscussed, such as the radio-frequency amplifier and the intermediate-frequency amplifier.

Second only to obtaining a picture of high quality, the aim of television receiver design is simplicity. The controls required are made few in number and simple to operate, and the construction of the equipment rugged and permanent. Only in this way can the television receiver be a truly satisfactory home instrument.

This makes it necessary to place the burden of producing a nearly perfect video signal upon the pickup and transmitting equipment, in order to free the receiver from the necessity of correcting controls such as are required for shading, etc. Therefore the sole duty of the receiver is to convert the corrected video signal into a picture.

In spite of this simplification, the television receiver is much more complicated than a broadcast receiver. Even the smallest table model reproducer requires as many as a dozen vacuum tubes, while a larger, more complete console receiver may require more than thirty. Consequently the design and manufacture of a television outfit suitable for use in the average home is no mean engineering feat.

17.1. Elements of the Television Receiver. All but the least expensive commercial television receivers are equipped to handle the accompanying sound as well as the picture. Thus, except for a common input stage and detector, the instrument really consists of two receivers: one for sound, the other for picture reproduction.

As was pointed out in Chapter 7, either a superheterodyne receiver or one involving straight radio-frequency amplification can be used. Because of the complexity of the latter, where reception on more than one television channel is desired, the former is rapidly gaining favor to the

exclusion of the other. Therefore only the superheterodyne receiver will be discussed in this chapter.

The elements of a receiver equipped for sound and picture are shown in block diagram in Fig. 17.1. Between each two of the units in this diagram, the form of the signal is indicated for convenience in tracing out the functions of the various parts. The major subdivisions of the receiver can be listed as follows:

1. Receiving antenna.
2. Ultra-high-frequency input circuit. One or more stages of radio-frequency amplification may be included in this circuit.
3. First detector and oscillator.
4. Intermediate amplifier.
5. Second detector.
6. Video amplifier and d-c reinsertion circuits.
7. Kinescope.
8. Synchronizing selector circuits.
9. Deflection generator.
10. Audio system.
11. Voltage supplies.

17.2. The Receiving Antenna. If the signal brought into the television receiver by the input transmission line has a low signal-to-noise ratio, or if it contains reflections and interference, even the most elaborate equipment cannot reproduce a good picture. On the other hand, a strong input signal with no interference and a high signal-to-noise ratio will give good results with even a very modest television receiver.

The two primary factors which determine the signal input are the antenna and the field strength of the direct radiation from the transmitter in the immediate vicinity of the antenna. The quasi-optical properties of ultra-high-frequency electromagnetic waves result in pronounced shadow and reflection effects from relatively small obstacles, such as buildings, gas or water tanks, and bridges. Consequently the direct radiation field strengths may be radically different at points separated by only a few feet. Because of this, the location of the antenna must be selected with care, even in a locale where the average field strength is high. The considerations involved in the design of the antenna itself are different from those of the broadcast receiving antenna, which is usually short compared with the received wavelength, because the former is in general one-half or more wavelengths long and constitutes in itself a tuned element.

The selection of a suitable location for a receiving antenna is usually made by a series of tests with a television receiver equipped with a port-

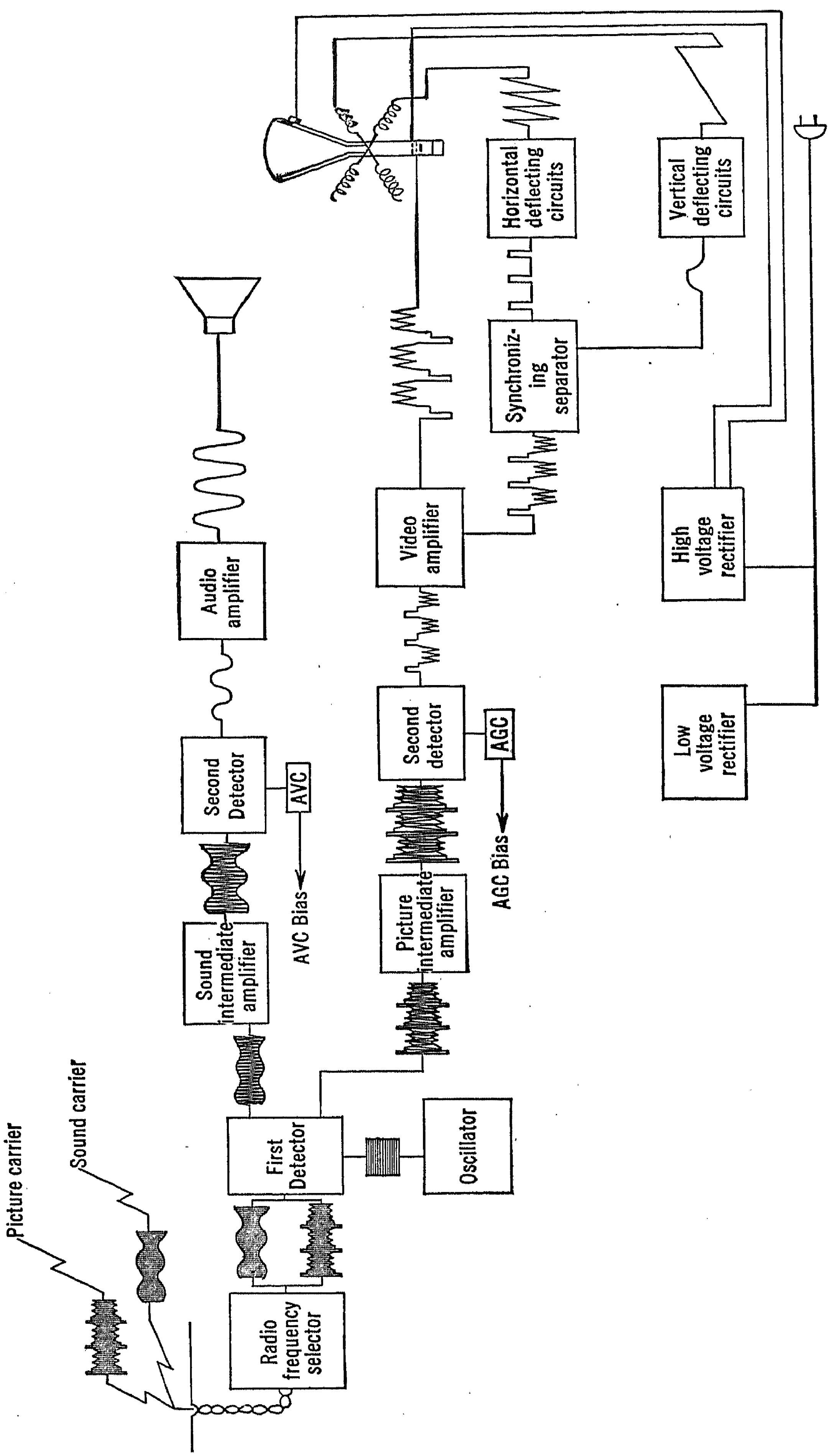


FIG. 17.1.—Block Diagram of Superheterodyne Receiver Showing Wave Forms at Various Points.

able antenna. Starting with well-elevated positions, preferably those from which the transmitting antenna can be seen, the antenna is moved until a satisfactory picture is obtained. The permanent antenna is then located in this position. Where there are several equally satisfactory possibilities, the choice is merely one of convenience.

The most important factor influencing the desirability of a given position is the presence or absence of reflections. These reflections have the effect of causing a blurring of the image, or even the appearance of multiple images. This is due to the delay in time of arrival of the reflected signal as a consequence of the longer path it must travel. At first sight it might seem that the delay caused by a few hundred feet of additional

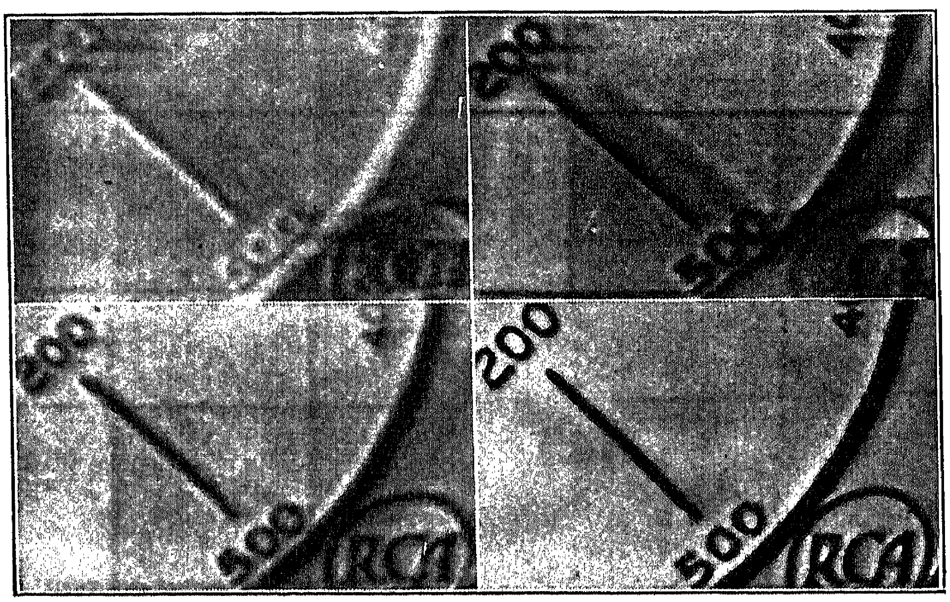


Fig. 17.2.—Effect of Reflections on Reproduced Picture (S. W. Seeley).

path length would be inconsequential. However, when it is remembered that the linear velocity of the spot across a 12-inch Kinescope screen is in the neighborhood of 130,000 inches per second, it will be realized that the spot moves 0.00014 inch per foot of signal travel. Consequently a reflection path which is only 300 feet longer than the direct path will result in a secondary image displaced by 0.04 inch from the wanted image, seriously blurring it. The effect of reflections is clearly visible in the reproduced picture* shown in Fig. 17.2.

Reflections, of course, occur from any conducting surface such as metal roofs and tanks. However, they are not restricted to conductors alone, for the difference in dielectric constant between such building materials as stone, brick, concrete, or even ordinary soil and the surrounding air is sufficient to produce strong reflections at small glancing angles.

* See Seeley, reference 16.

Thus any large buildings between the transmitting and receiving antennas are potential sources of this type of interference.

Interference by reflections is avoided in part by properly positioning the antenna, as has already been pointed out, and in part by making use of the directional properties of the ultra-short-wave antenna.

The most frequently used television receiving antenna is the half-wave doublet. This type of antenna has already been introduced in the preceding chapter. It consists of a straight wire whose length is equal to one-half a wavelength. For ordinary receiver work the antenna is divided at the center and feeds a balanced transmission line. When coupled in this way it has a radiation resistance of approximately 72 ohms. Such an

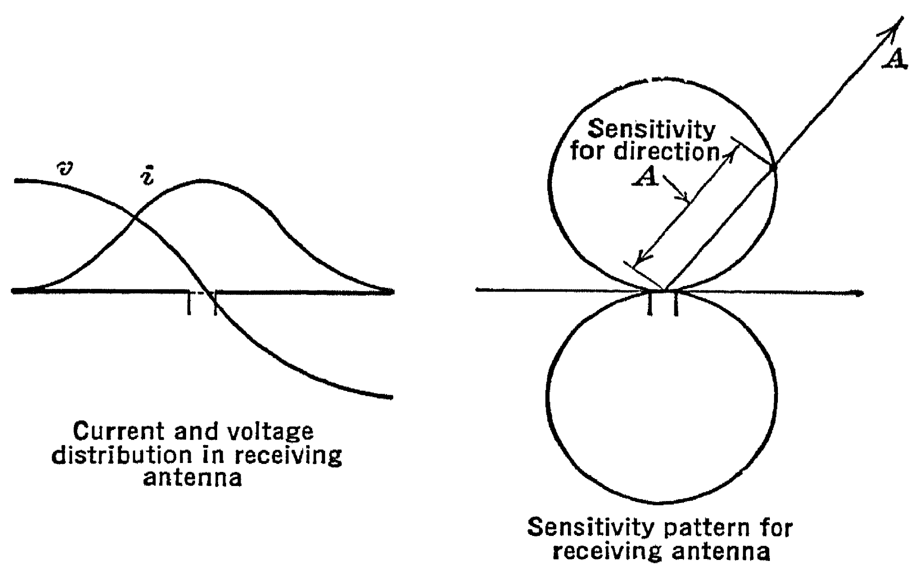


Fig. 17.3.—Electrical Behavior and Directivity of a Half-Wave Receiving Dipole.

antenna is resonant at one frequency only. However, when taken in connection with its transmission line and the input circuit of the receiver, its response can be made fairly uniform over the entire band contemplated for television broadcasting. Fig. 17.3 illustrates the current distribution in this type of antenna, together with its directional properties. In the diagram indicating the directional properties the distance from the center of the pattern to a point on the loop represents the sensitivity in the direction connecting these points.

A receiver located in the border regions of the service area may require an antenna with more sensitivity than the half-wave dipole for satisfactory performance. The added sensitivity is most frequently obtained by means of a reflector in conjunction with the antenna. The reflector consists of a half-wave conductor placed parallel to, and at a quarter wavelength distance behind, the antenna.

The sensitivity and directional selectivity of a linear antenna increases

if it is several multiples of a half wavelength long. The response patterns of a number of long antennas are shown in Fig. 17.4. Even greater directivity can be obtained with the "Vee beam" and rhombic antennas illustrated in Fig. 17.5.

The antenna is usually located at some distance from the receiver, and the signal must be fed to the receiver through a transmission line. In most practical cases a suitably designed twisted pair will suffice. While the losses in such a line are as high as 1.5 to 2 db per wavelength of line, and prevent the economical use of more than 40 or 50 feet, it is relatively insensitive to interference when balanced and is very easy to install. The line should be matched to the impedance of the antenna, but because of the wide frequency range to be covered, and because of a number of factors which cannot be accurately evaluated, an exact match is not warranted. A line having an impedance of about 100 ohms is found to operate satisfactorily into a simple doublet whose resonance impedance is 72 ohms.

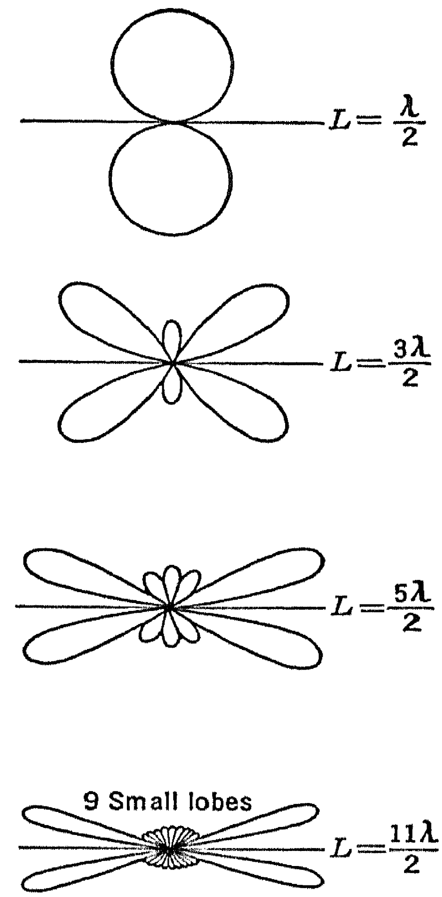


Fig. 17.4.—Directional Characteristics of Linear Antennas.

For localities where the field strength is low, or where very long transmission lines are required, a more efficient line may be necessary. The two types commonly employed under these conditions are the concentric cable and the open-wire transmission line. The second type of

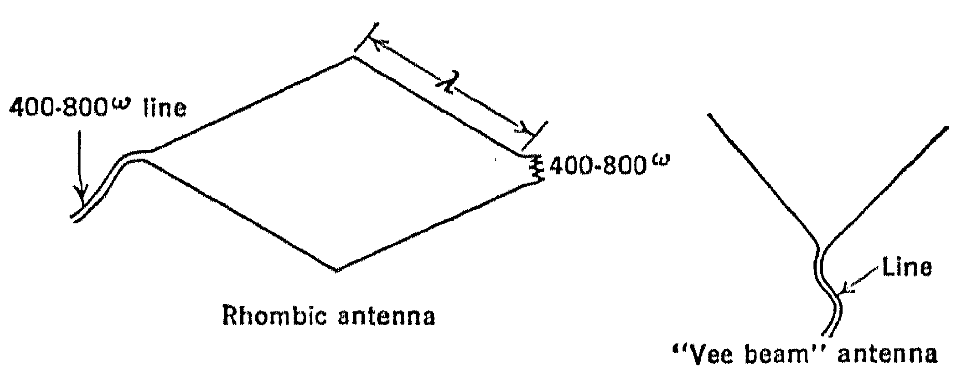


Fig. 17.5.—Some Highly Directive Antennas.

line is particularly advantageous when the radiation resistance of the antenna is high, because its surge impedance, which is given by $Z_s = 276 \log (2D/d)$, where D is the distance between wires and d the diameter

of the wire, can easily be made high by spacing the wires properly. Without any special precautions, the loss in an open-wire transmission line will be roughly 0.1 db per wavelength of line.

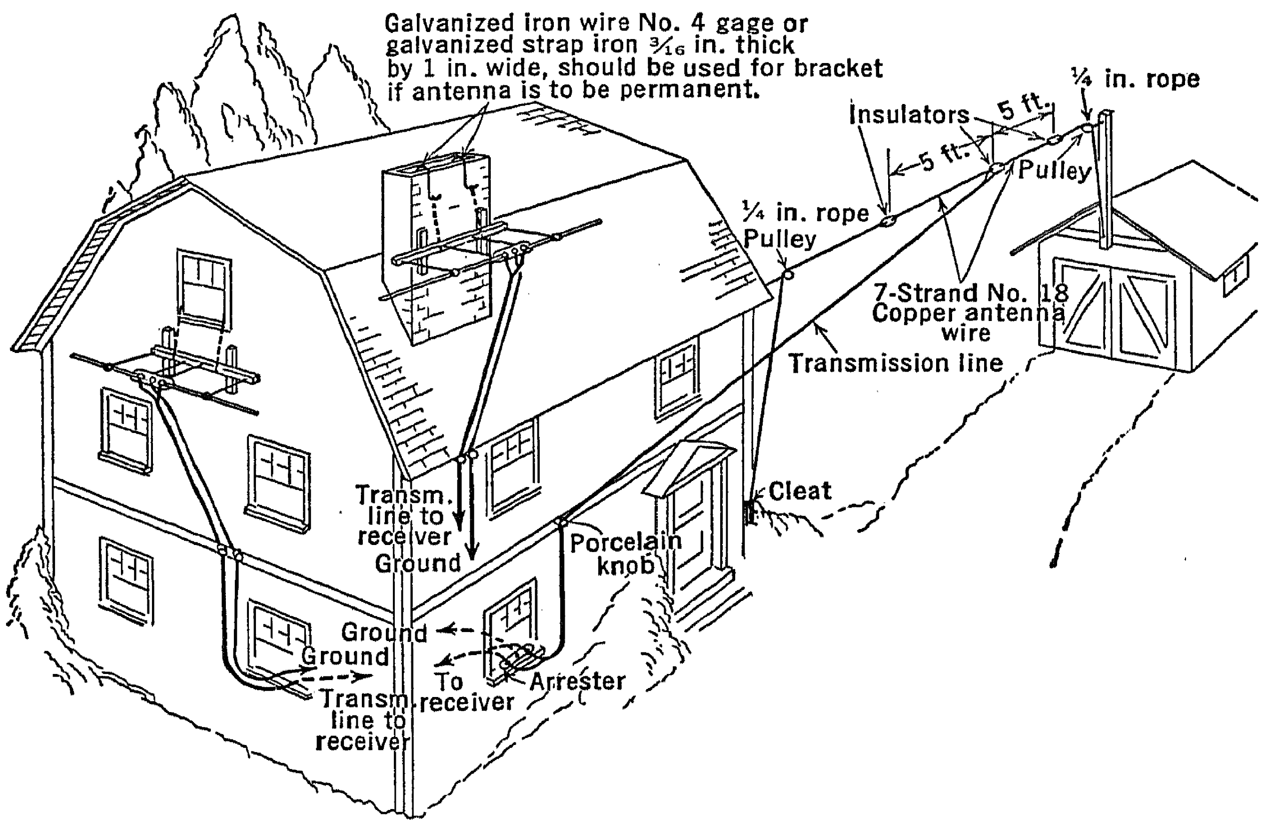


Fig. 17.6.—Practical Television Receiving Antenna Installations.

The appearance of a practical construction of a television doublet antenna, together with several typical installations, is shown in Fig. 17.6.

17.3. The Input Circuit and First Detector. The input circuit couples the antenna with the first detector or mixer. As such it must match the transmission line from the antenna in order to avoid reflections which

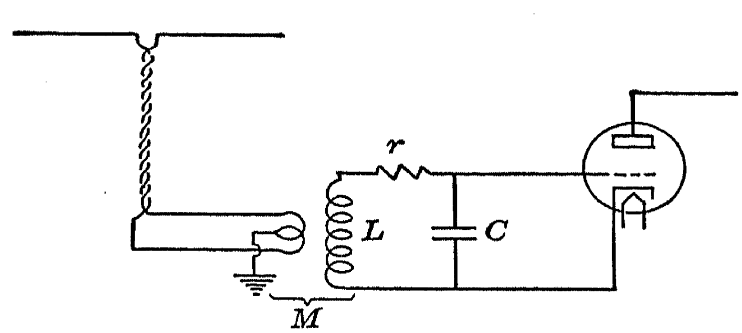


Fig. 17.7.—Simple Input Circuit.

would lead to loss of detail. One of the simplest coupling circuits which can be used between the antenna and transmission line and the detector or r-f amplifier tube is a transformer with a tuned secondary and a small untuned primary. This is illustrated in Fig. 17.7. The inductance of the

secondary is L, and it is shown shunted by a capacity C composed of the variable tuning condenser C_v and a minimum capacitance C_m. This minimum capacity is the sum of the input capacity of the first tube, the wiring capacity, and the residual capacity of the tuning condenser. The series resistance r represents the losses in the circuit. It may be increased by additional series or shunt resistance if the bandwidth requirements make this necessary. If the mutual inductance between primary and secondary is M, and the inductance of the primary is assumed to be negligible compared with that of the tuned secondary, the effective impedance of the primary is:

$$Z_p = \frac{(2\pi f M)^2}{Z_s}. \tag{17.1}$$

This impedance must match that of the transmission line if reflections are to be avoided, and therefore must be considered as a fixed quantity.

When a signal voltage e_1 is applied to the primary, the voltage e_2 across the secondary is given by:

$$e_2 = |2\pi f M i_p| = \left|\frac{e_1 Z_s}{2\pi f M}\right|,$$

and, substituting from equation 17.1, this becomes:

$$e_2 = e_1 \left|\sqrt{\frac{Z_s}{Z_p}}\right|. \tag{17.2}$$

Therefore the voltage transfer $T = e_2/e_1$ is proportional to the square root of the impedance of the tuned circuit.

The secondary circuit is a resonant network whose impedance varies with frequency. In order to maintain adequate sensitivity over the range of frequencies covered by the sidebands, the circuit constants must be so chosen that the minimum impedance of this circuit is at least 70 per cent of the maximum impedance. The circuit meeting this requirement can be determined as follows: Define the Q of the circuit in the ordinary way, namely:

$$Q = \frac{2\pi f_r L}{r} = \frac{1}{2\pi f_r C r}, \tag{17.3}$$

where the resonant frequency f_r is given by:

$$f_r = \frac{1}{2\pi\sqrt{LC}}.$$

The frequency band $2\Delta f$, over which the impedance is greater than 0.7 of resonant impedance, is given by:

$$\frac{\Delta f}{f_r} = \frac{1}{Q}. \tag{17.4}$$

The resonant impedance of the secondary circuit in terms of Q is:

$$Z_s = \frac{Q}{2\pi f_r C}.$$

From equations 17.2, 17.3, and 17.4, it follows that the voltage transfer will be related to the circuit constants, and the frequency band, as follows:

$$T \sim \frac{1}{\sqrt{2\pi \Delta f C}}. \tag{17.5}$$

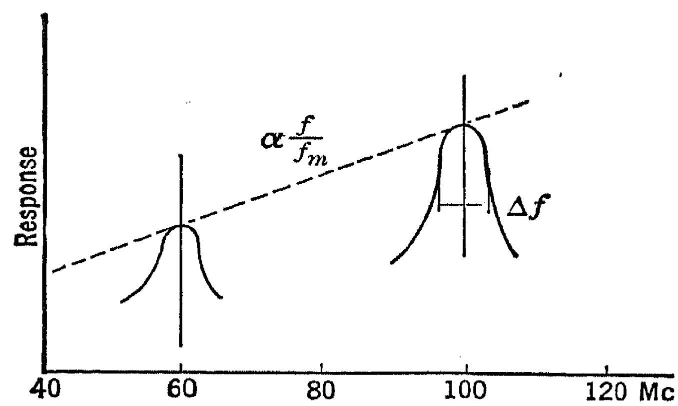


FIG. 17.8.—Response of Simple Input Circuit.

Equation 17.5 shows that the maximum transfer which can be obtained for a given frequency band is limited by the shunt capacity. This has a minimum value C_m corresponding to the maximum frequency f_m of the region over which the circuit can be tuned. In terms of this capacity, equation 17.5 becomes:

$$T \sim \frac{1}{\sqrt{2\pi \Delta f C_m}} \frac{f_r}{f_m}.$$

This makes evident the need for small tube and wiring capacities.

It will be noticed that the transfer varies linearly with the frequency to which the circuit is tuned. If the receiver is to be useful over the entire television broadcast band from 40 to 100 megacycles this variation is more than two to one. Clearly such a variation is undesirable, and though it can be improved somewhat by judiciously selecting the resonant qualities of the antenna and by permitting a mismatch to the

transmission line at some frequencies, it constitutes a limitation to the usefulness of the circuit. Furthermore, since the mutual inductance must be varied as the circuit is tuned, it does not lend itself readily to single knob tuning. Finally, it is possible to obtain a higher voltage transfer with other more elaborate nets. However, for purposes of discussion its basic simplicity makes it suitable for illustrating the principles of input coupling.

Substituting a transformer having a tuned primary as well as a tuned secondary in place of the simpler circuit shown in Fig. 17.7 leads to a voltage gain which is nearly twice as great for the same bandwidth. This

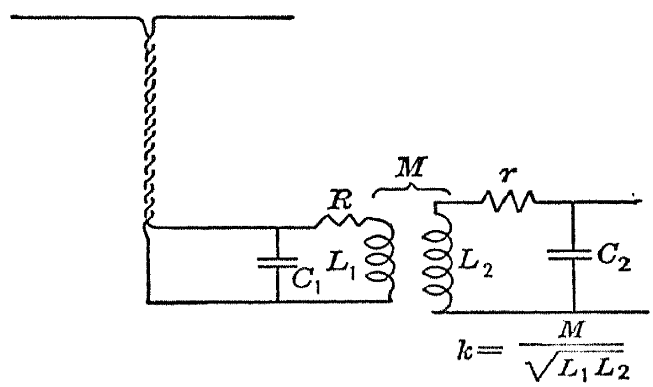


FIG. 17.9.—Input Circuit Employing Tuned Transformer.

modification is shown in Fig. 17.9. The gain, near resonance, of this circuit is given by the following familiar relation:

$$T = \left[\frac{k^2 \dfrac{L_2}{L_1}}{16\left(\dfrac{\Delta f}{f}\right)^4 + 4\left[\dfrac{1}{Q_1^2} + \dfrac{1}{Q_2^2} - 2k^2\right]\left(\dfrac{\Delta f}{f}\right)^2 + \left[\dfrac{1}{Q_1 Q_2} + k^2\right]^2} \right]^{\frac{1}{2}}.$$

In this expression f is the resonant frequency and is equal to:

$$f = \frac{1}{2\pi\sqrt{L_1 C_1}} = \frac{1}{2\pi\sqrt{L_2 C_2}},$$

while Δf is the difference between the frequency under consideration and resonance. Q_1 and Q_2 are defined in the ordinary way, where

$$Q_1 = \frac{2\pi f_r L_1}{R} \text{ and } Q_2 = \frac{2\pi f_r L_2}{r},$$

and finally k is the coefficient of coupling between the two coils. The character of the gain curve for this circuit is shown in Fig. 17.10.

The insertion of a tuned link between an untuned antenna coil and

the tuned secondary circuit results in still further improvement, giving a greater useful bandwidth for a given selectivity. This circuit is shown in Fig. 17.11, and the form of its response curve in Fig. 17.12.

The primary of every transformer-coupled input circuit has a center tap to ground. This is necessary to balance the transmission line. For example, if one side were grounded instead, any voltage induced by in-

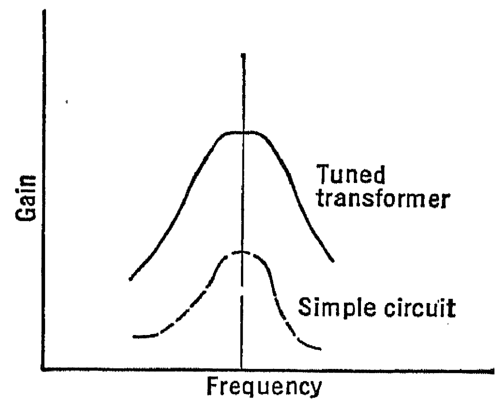


Fig. 17.10.—Comparative Response of Simple and Tuned Transformer Input Circuits.

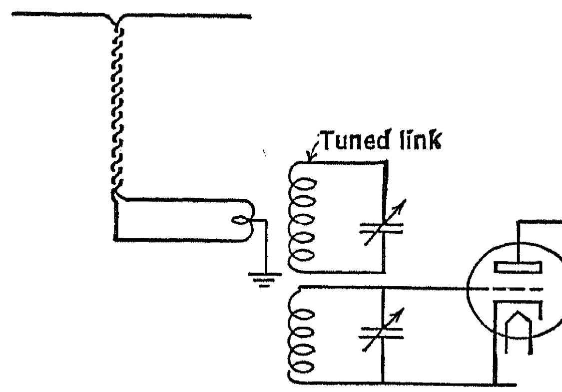


Fig. 17.11.—Input Circuit with Tuned Link Coupling.

terference in the two transmission line wires would cause a current to flow in the primary of the coupling transformer, and so into the receiver. On the other hand, with the center of the coil grounded, the net current flow is zero unless different voltages are induced in the two wires, which is avoided by the use of a close-spaced, twisted pair.

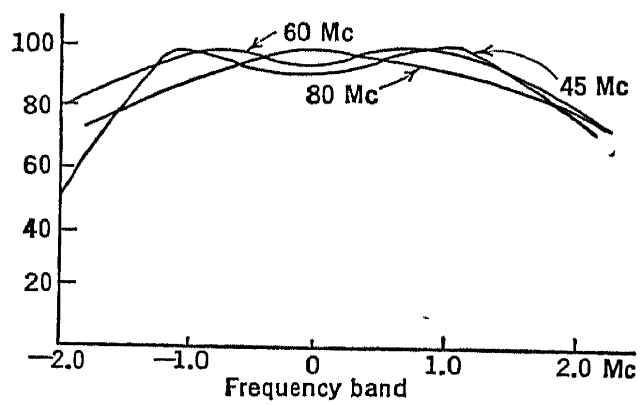


Fig. 17.12.—Response Characteristics of Input Circuit Shown in Fig. 17.11.

The voltage generated across the secondary of the input transformer is applied to the grid of the first tube, which may be the mixer or a radio-frequency amplifier.

Three possible advantages can be obtained by the insertion of a radio-frequency stage. First it causes a smaller decrease in signal-to-noise ratio than when the coupling circuit feeds the mixer directly. Second it in-

creases the selectivity of the ultra-high-frequency input circuit through the addition of a further tuned element, and thereby reduces the tendency toward image interference in the intermediate amplifier. Finally it acts as a buffer to prevent the output of the local oscillator from reaching the antenna and being radiated.

Against these advantages is the fact that the gain of such a stage is very low compared with that obtained from a stage of intermediate frequency amplification. Furthermore, it considerably increases the complication of the input circuit by requiring the addition of one or more tuned circuits.

Without a radio-frequency stage the tuned coupling circuits can, in practice, be made sufficiently selective to avoid serious image interference except for extremely strong interfering signals. This selectivity is also sufficient to reduce radiation from the local oscillator to a low level. The question whether or not a radio-frequency stage is required, unless dictated in certain instances by production considerations, is primarily determined by whether the improvement in signal-to-noise ratio warrants its inclusion.

In Table 14.1, a number of tubes suitable for television use were listed, giving their transconductance and their input and output capacitances. Except for the RCA 1852 and 1853, none of these tubes is suitable for wide-band ultra-high-frequency amplification because the relatively low mutual conductance and necessarily low impedance of the external plate circuit make the possible gain less than 2. The RCA 1852 has a transconductance of 9000 micromhos and, working into an output impedance of 1000 ohms, which is about the maximum possible for this frequency and bandwidth, can give a gain of 9. Under similar conditions the RCA 1853 will have a gain of 5. These two tubes, then, are satisfactory.

Assuming that a radio-frequency stage employing an RCA 1852 is used, and taking the effective resistance of a grid input circuit of the type shown in Fig. 17.11 as 2000 ohms, the total noise referred to the plate is:

$$\begin{aligned}\overline{E_N} &= \sqrt{K_1\Delta f}\ \sqrt{R_G + R_N}\cdot g_m Z_p \\ &= \sqrt{K_1\Delta f}\cdot g_m Z_p \sqrt{2000 + 520},\end{aligned}$$

and if the input signal is e_s, the signal-to-noise ratio T at the plate is:

$$T = \frac{e_s g_m Z_p}{\sqrt{K_1 \Delta f} g_m Z_p \sqrt{2500}} = \frac{0.02 e_s}{\sqrt{K_1 \Delta f}}.$$

In these relations Δf is the frequency bandwidth; K_1, the coefficient of thermal agitation at room temperature; and R_N, the equivalent noise resistance of the tube.

The alternative is to operate the antenna coupling circuit directly into the first detector. The most satisfactory tube of those available at present for a mixer is the RCA 1852. This tube has a conversion conductance g_c of about 3000 micromhos. Assuming the circuit conditions of the previous example, the noise can be estimated as follows: The tube noise will appear in the plate output as before, but, because the local oscillator cuts off the plate current for part of the cycle, it is materially reduced. The actual value varies for different operating conditions, but may, within the accuracy of the present estimate, be assumed to be one-half the normal noise. Therefore the noise component from the tube alone is:

$$\overline{E_{TN}^2} = \frac{1}{2}K_1\Delta f \cdot R_N g_m^2 Z_p^2.$$

Added to this, however, is the thermal noise of the coupling circuit, which is:

$$\overline{E_{cN}^2} = K_1\Delta f \cdot R_G g_c^2 Z_p^2.$$

Consequently the total root mean square noise is:

$$\overline{E_N} = \sqrt{\overline{E_{TN}^2} + \overline{E_{cN}^2}} = \sqrt{K_1\Delta f} \cdot \sqrt{R_G + \frac{R_N g_m^2}{2g_c^2}} \cdot g_c Z_p$$
$$\sim \sqrt{K_1\Delta f} \cdot g_c Z_p \sqrt{4500},$$

and the signal-to-noise ratio is given by

$$S \sim \frac{e_s g_c Z_p}{\sqrt{K_1\Delta f} \cdot g_c Z_p \sqrt{4500}} = 0.015 \frac{e_s}{\sqrt{K_1\Delta f}}.$$

The stage of radio-frequency amplification therefore increases the signal-to-noise ratio by less than a factor of 2 and consequently in this instance is not justified. However, when a mixer tube having a poorer signal-to-noise ratio than the RCA 1852 serves as first detector, it may be advisable.

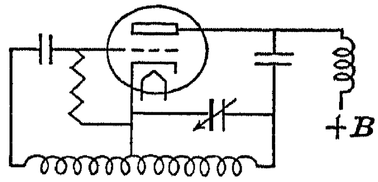


FIG. 17.13.—Local Oscillator of First Detector.

From the standpoint of tracking and of frequency stability, it has been found in television practice that more satisfactory results are obtained if a separate oscillator is used with the first detector than if a single tube performs the functions of both mixer and oscillator. The oscillator is usually of more or less conventional form, the tuned plate feedback oscillator, illustrated in Fig. 17.13, being found very satisfactory. The tube for this purpose should have a fairly high transconductance to capacity ratio. Such tube types as the RCA 955, 6J5 or

6K8 have the desired characteristics. The oscillator can be coupled to the grid of the mixer by inductance, capacitance, or conductance, or a combination of these methods. About 3 volts are required on the grid of an RCA 1852 to take full advantage of its high conversion conductance. The frequency of the local oscillator is, for practical reasons such as that of reducing video band to intermediate carrier ratio and decreasing likelihood of image interference, usually chosen higher than the frequency of the signal. Since the intermediate frequency is usually in the neighborhood of 10 to 15 megacycles, the oscillator must be variable over a range of frequencies of 50 to 125 megacycles to cover the television broadcast region.

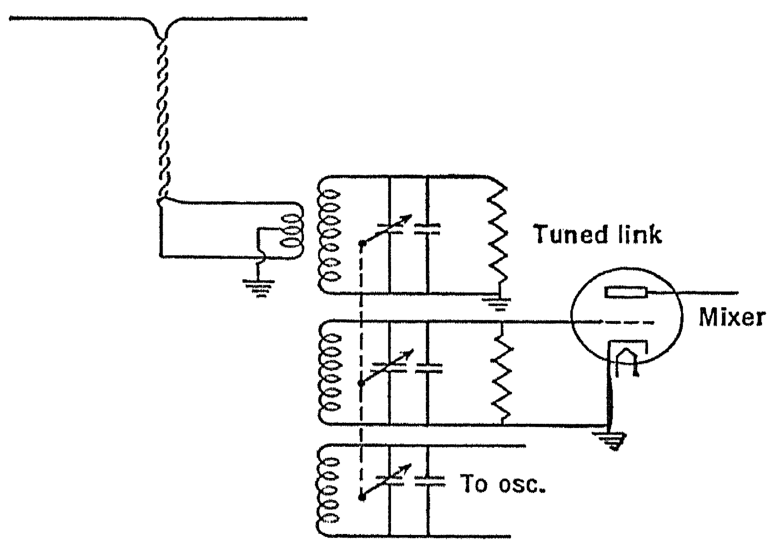


Fig. 17.14.—Complete Input Circuit.

The circuit diagram of the input circuit of an RCA experimental receiver is shown in Fig. 17.14. The antenna transmission line is terminated in a coil which is inductively coupled to a tuned link circuit. This link circuit includes a small trimmer condenser and one member of a gang tuning condenser. The mixer tube input circuit consists of similar elements. A tuned plate feedback oscillator is controlled by the third member of the gang condenser. The three tuned circuits are mutually coupled by the conductance of the common rotor shaft of the tuning condensers. A selectivity adequate to avoid difficulties resulting from image interference in the intermediate amplifier circuit is obtained without the need of additional filtering such as might be afforded by a radio-frequency stage. A photograph of the assembled input circuit used by an RCA experimental television receiver is shown in Fig. 17.15.

17.4. Television Receiving Tubes. The importance of high transconductance and low input and output capacitance has been made evident

in the discussion of the video amplifier and the transmitter. Similar considerations are seen to be involved in tubes employed for ultra-high-frequency amplification in the receiver. In general it may be said that the gain which can be obtained from a stage of amplification depends upon the transconductance of the tube and the impedance of the coupling circuit. The latter, for a given bandwidth, is inversely proportional to the sum of the plate and grid capacities, where a two-terminal net is involved; and in the limiting case of a four-terminal coupling circuit, to the inverse

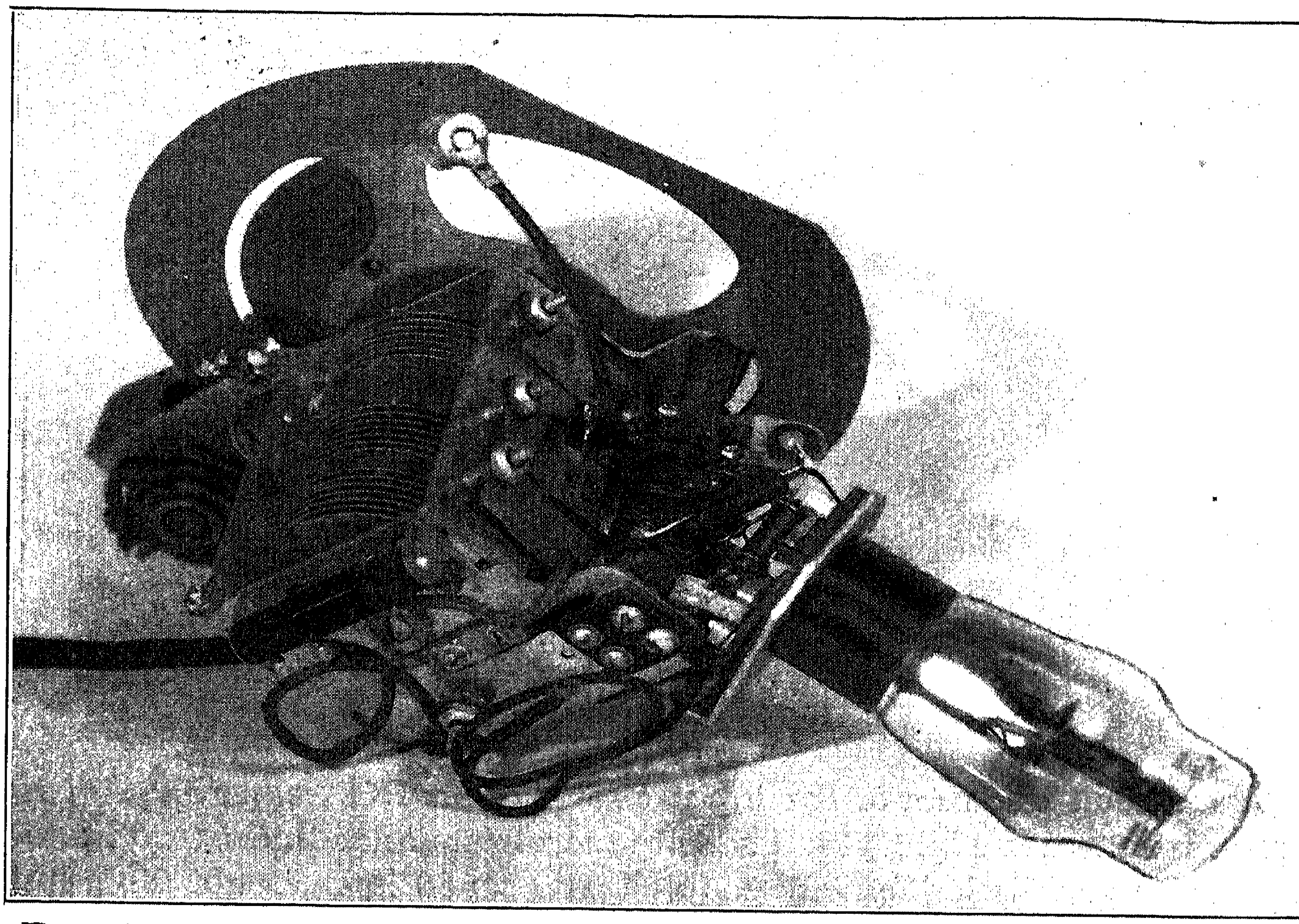


FIG. 17.15.—Photograph of Input Assembly of an RCA Experimental Receiver.

geometric mean of the two capacities. For wide-band ultra-high-frequency amplification, where the interstage coupling impedance is limited to a few hundred ohms, the gain per stage is close to unity or even fractional if ordinary tubes having transconductances of 1000 to 2000 micromhos are employed. Even at the frequencies used for television intermediate amplifiers the gain per stage with conventional tubes is only 2 to 5. This has made necessary the design of special tubes for such service.

The transconductance of a tube can be increased merely by increasing the area of the cathode while maintaining the grid and plate spacing constant. This, however, increases the capacities proportionately, so

that the ratio $g_m/\sqrt{C_gC_p}$ remains constant, and there is no improvement in the tube for the purpose in question.

An alternative is to decrease the grid-to-cathode spacing. The geometric input capacity is inversely proportional to this spacing, while the transconductance varies approximately with the inverse square. Therefore if the spacing is reduced to one-half its original value the ratio $g_m/\sqrt{C_gC_p}$ is more than doubled since the output capacity is but little affected by the change. This procedure has been used in the design* of the RCA 1852 and 1853 and has led to tubes having transconductances of 9000 and 5000 micromhos, respectively, while the input capacities are only 13.5 and 9.3 micro-microfarads under normal operating conditions.

The input capacitance can be divided into two parts: one, the geometric capacity between grid and cathode (and screen, etc.) which is usually known as the cold capacity, and in the RCA 1852 constitutes about 11 micro-microfarads of the total capacity; the other, the capacity caused by the displacement current induced by the motion of the electrons in the vicinity of the grid. The origin of this portion can be explained by considering the instantaneous cloud of electrons moving towards the grid between cathode and grid, together with that moving away between anode and grid. This is illustrated in Fig. 17.16. If the time taken by the electrons in moving from cathode to anode is short compared with the period of the voltage applied to the grid, it will be evident that the net displacement current induced in the grid will be in quadrature with the voltage, and the effect will be that of an increase in capacity. This component of capacity which is contributed by the electron flow depends upon the current through the tube. It is therefore a function of grid bias, and, where the tube is used in a bias-controlled variable-gain stage, due account must be taken of the change in capacities.

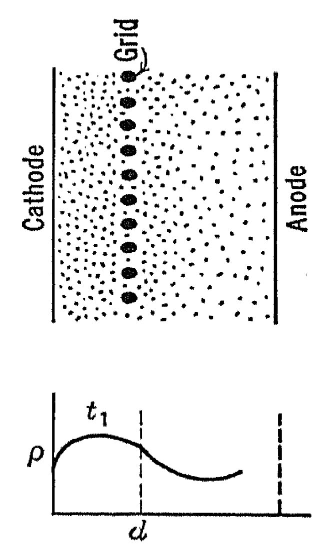


FIG. 17.16.—Electron Distribution in Triode.

At very high frequencies the input capacity alone does not determine the effect of the tube on the input circuit, inasmuch as the conductive or real component of the input admittance may be high enough to be important. The transit time, at these frequencies, may become comparable with the period of oscillation, so that the change in electron density in the vicinity of the grid lags behind the change in voltage. This is so because the electrons near the cathode, where the control of the grid really occurs, do not reach the grid until its potential has considerably altered.

* See Kauzmann, reference 20.

Consequently the displacement current will have a component in phase with the grid voltage. This conductance is known as transit time loading. A theoretical analysis shows that the transit time conductance increases with the square of the frequency, while the capacitative effect of the displacement current decreases directly with the frequency.

Transit time loading constitutes only a portion of the conductive component of the input admittance. The inductance of the cathode lead which is common to the input and output circuits has also been shown to constitute an input loading. Since it varies with the square of the frequency it cannot be directly distinguished from transit time loading. Finally, ion current and electron current from the grid may constitute a source of loading, but in a well-designed tube this type of conduction can be made negligible in comparison with the previously mentioned factors at high frequency. At 40 megacycles the transit time loading in the RCA 1853 is from 3 to 12 micromhos out of a total of 230 micromhos. The ion current and grid emission current is only of the order of a microampere. This current, since it is more or less independent of frequency, does, however, play a role in limiting the maximum safe external grid resistance that can be used. If the grid resistor is made too large the direct current through it will make the grid go positive. With a 160-ohm cathode bias resistor, the maximum grid resistor which can be safely used is 250,000 ohms. It can be increased to 500,000 ohms if the bias resistor is increased to 190 ohms and the screen voltage is obtained through 30,000 ohms from a 300-volt supply.

It should be mentioned in connection with these special television tubes that, owing to the extremely close spacing between the cathode and grid, they are very sensitive to changes in contact potential and mechanical displacement of the parts. For this reason it is advisable to obtain the bias from a resistor in the cathode lead rather than directly from a voltage supply.

The special tubes just described were designed to meet the demands of the television intermediate amplifier, as well as for use in the ultra-high-frequency and video circuits.

17.5. The Intermediate Amplifier. By far the greater part of the amplification in a superheterodyne television receiver takes place in the intermediate amplifier. Here also the sound and picture signals are separated from one another and the video response characteristic is given the desired shape.

The factors influencing the choice of the intermediate frequency are based on the availability of tubes and circuits which will give the required amplification over the entire channel and a consideration of interference due either to direct pickup from local transmitters or to image response. A high intermediate frequency is desirable because it minimizes

image interference and reduces the ratio between bandwidth and carrier frequency. On the other hand, frequencies above 15 or 20 megacycles introduce difficulties both in the matter of tubes and circuits. Finally,

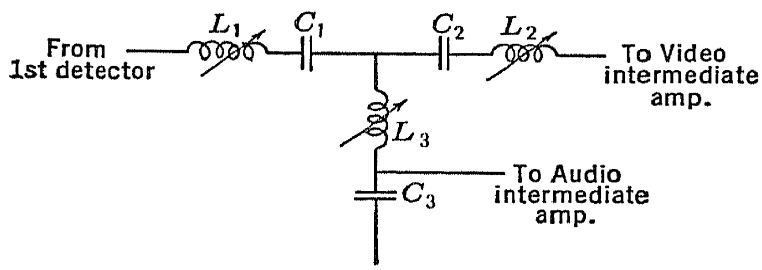


FIG. 17.17.—Selector Separating Video and Audio Intermediate-Frequency Signals.

there exist amateur bands at 14 and at 7 megacycles. These considerations indicate that the frequencies between 13 and 8 megacycles are suitable. The RMA has selected 12.75 megacycles as a tentative standard for the intermediate frequency.

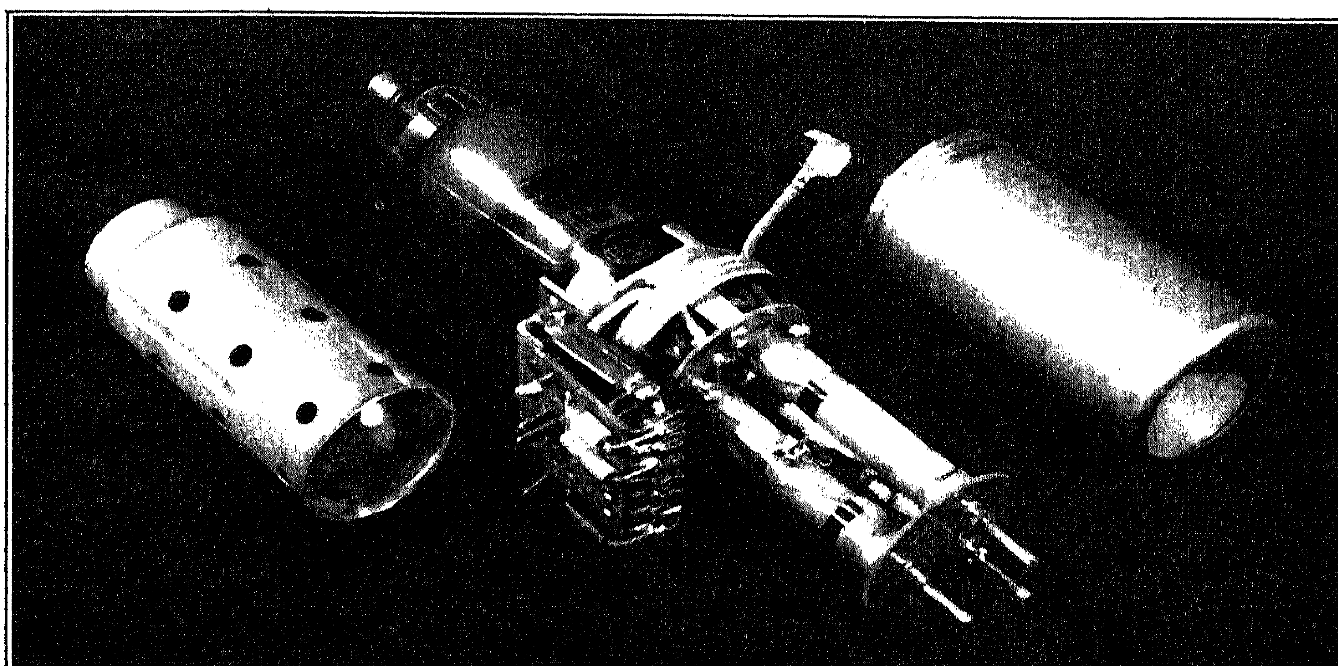


FIG. 17.18.—Selector and Intermediate-Frequency Coupling Assembly Used in an RCA Experimental Television Receiver.

Since the local oscillator which supplies the converter is usually made to have a higher frequency than that of the signal, the audio intermediate carrier will lie at the lower end of the intermediate-frequency spectrum, while the video carrier will be at the top.

The separation of sound and picture is usually made at the output of the first detector. There are many circuits which may be used for this purpose. One which has experimentally been found to be very successful is shown in Fig. 17.17. As far as the video signal is concerned, it is a T

band-pass section with the branches containing L_1C_1 and L_2C_2 tuned to the video carrier. L_3C_3 is tuned to the audio carrier, so that a large audio signal appears across C_3. A practical embodiment of this transformer is shown in Fig. 17.18. The coils are made with magnetite cores, and tuning is done by adjusting these, rather than with variable condensers, to minimize the capacity necessary in the circuit. This gives a high L/C ratio which, in turn, permits the most gain for a given bandwidth. Fig. 17.19 shows the audio and video selectivity.

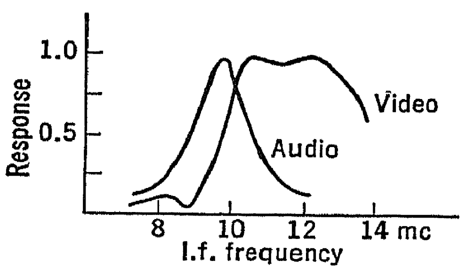


FIG. 17.19.—Response of Circuit Shown in Fig. 17.17 to Audio and Video Intermediate-Frequency Signals.

The simplest interstage coupling for an intermediate amplifier is a two-terminal band-pass filter, consisting merely of a tuned circuit such as is illustrated in Fig. 17.20. For purposes of calculation, its equivalent circuit is also included. The capacity C includes the input capacity of the first tube and wiring. The inductance L

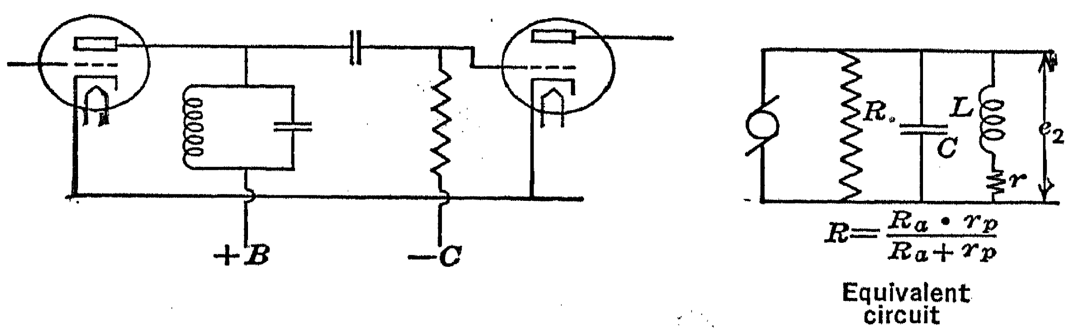


FIG. 17.20.—Simple Intermediate-Frequency Coupling.

is assumed to have a resistive component r, which, however, is small in a well-designed coil. The gain G of this stage is:

$$G = \frac{e_2}{e_1} = g_m \frac{RZ}{R + Z},$$

$$Z = \frac{r + 2\pi jfL}{1 - 4\pi^2 f^2 LC + 2\pi jfrC}.$$

Under the assumption that the resistive loss r of the coil is small, the magnitude of the gain becomes:

$$G = g_m \frac{R}{\sqrt{1 + \dfrac{(1 - 4\pi^2 f^2 LC)^2 R^2}{4\pi^2 f^2 L^2}}}.$$

Resonance occurs when $f^2 = 1/(4\pi^2 LC)$, and at this frequency the gain is merely

$$G = g_m R.$$

At either side of resonance the gain decreases. At some arbitrary frequency $(f + \Delta f)$ the gain may be written as:

$$G = \frac{g_m R}{A},$$

where

$$A = \sqrt{1 + \frac{(1 - 4\pi^2(f + \Delta f)^2 LC)^2 R^2}{4\pi^2(f + \Delta f)^2 L^2}}$$

$$A \cong \sqrt{1 + 16\pi^2 C^2 R^2 \Delta f^2}.$$

If the total bandwidth to be amplified is taken as $2\Delta f$, and the uniformity of amplification requires that A cannot be greater than 1.4, the maximum value for R is limited to:

$$R = \frac{1}{4\pi \Delta f C}.$$

This in turn restricts the maximum gain to $G = g_m/(4\pi\Delta f C)$. The total capacity, assuming that an RCA 1853 is used, is:

$$C = \underset{\text{input}}{9.3} + \underset{\text{output}}{5.0} + \underset{\text{wiring}}{5.0} = 19.3\mu\mu\text{f}$$

and the mutual conductance 5000 micromhos. Therefore, if the bandwidth to be amplified is 4.25 megacycles, the maximum value permissible for R is about 2000 ohms and the gain of the stage is limited to approximately 10.

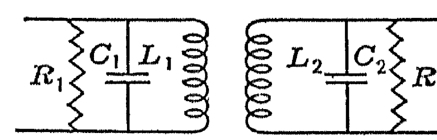


Fig. 17.21.—Tuned Transformer Used to Couple Intermediate-Frequency Amplifier Stages.

If a transformer with tuned primary and secondary replaces the resonant circuit,* both the selectivity and gain are increased, because it permits dividing and matching individually the capacities of the tubes. This type of circuit is illustrated in Fig. 17.21. The primary and secondary circuits are made resonant to a frequency near the edge of the band to be amplified in such a way that:

$$4\pi^2 f_r^2 L_1 C_1 = 4\pi^2 f_r^2 L_2 C_2 = 1.$$

The damping factor or Q at resonance of the two circuits will be:

$$Q_1 = \sqrt{\frac{L_1}{C_1}} \frac{1}{R_1},$$

$$Q_2 = \sqrt{\frac{L_2}{C_2}} \frac{1}{R_2}.$$

* See Cocking, reference 21.

These damping factors will be assumed to be alike for the primary and secondary circuits and equal to Q. Under these conditions the gain of the stage becomes:

$$G = g_m \frac{xQk\sqrt{R_1R_2}}{\sqrt{[(1 - x^2)^2 - x^2Q^2(1 - k^2) - x^4k^2]^2 + 4x^2Q^2[1 - x^2(1 - k^2)]^2}},$$

where

$$x = 2\pi f\sqrt{L_1C_1} = 2\pi f\sqrt{L_2C_2}; \text{ and } k = \frac{M}{\sqrt{L_1L_2}}.$$

This has its optimum value when the coupling coefficient k is related to the damping in the following way:

$$k = \frac{Q}{\sqrt{1 + Q^2}}.$$

Under these conditions the gain becomes:

$$G = g_m \frac{xQ^2\sqrt{1 + Q^2}\sqrt{R_1R_2}}{\sqrt{(1 - x^2)^2(1 + Q^2 - x^2)^2 + 4x^2Q^4(1 + Q^2)}}.$$

The general shape of this type of response is shown in Fig. 17.22. It consists of two maxima occurring at $x = 1$ and $x^2 = 1 + Q^2$ and at these points has the value:

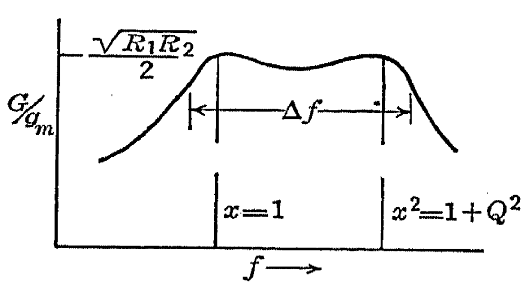


FIG. 17.22.—Response of Tuned Transformer Coupling.

$$G = g_m \frac{\sqrt{R_1R_2}}{2}.$$

It will be noticed that the gain varies with the values of R_1 and R_2. The maximum values of these resistances are in turn fixed by the values for Q which give the desired separation between the two peaks (more exactly edges of the band). For a given Q, R is dependent upon the capacities.

In practice it is difficult to obtain sufficient coupling between the primary and secondary of a conventional transformer. Therefore the circuit equivalent of the transformer is often used. This is shown in Fig. 17.23. In such a circuit the coupling is provided by the common inductance L_3.

This circuit and the transformer has a response which is approximately symmetrical about a midpoint. Since the sound channel closely adjoins the video channel on one side it is necessary to make this side very selective. On the other hand, the other side should have a tapering response. A response having the desired characteristic can be obtained by the

circuit shown in Fig. 17.24. The selectivity of such a rejector stage is given by Fig. 17.25. The high selectivity on the low-frequency side of the

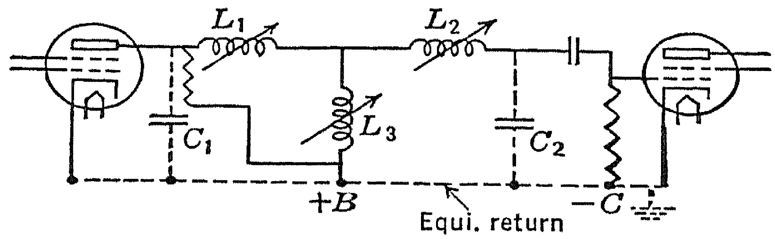


FIG. 17.23.—Intermediate-Frequency Coupling Circuit Equivalent to a Tuned Transformer.

response is obtained by tuning the series element of the net to the frequency of the audio intermediate frequency. Often more than one unsym-

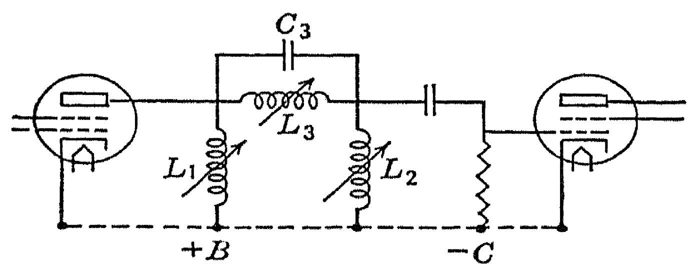


FIG. 17.24.—Interstage Rejector Coupling with Asymmetric Response.

metrical rejector stage is required to avoid completely interference in the picture due to the audio signal. The complete intermediate amplifier, for

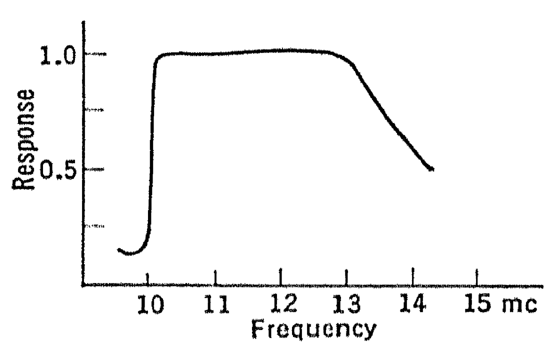


FIG. 17.25.—Response of Circuit Shown in Fig. 17.24.

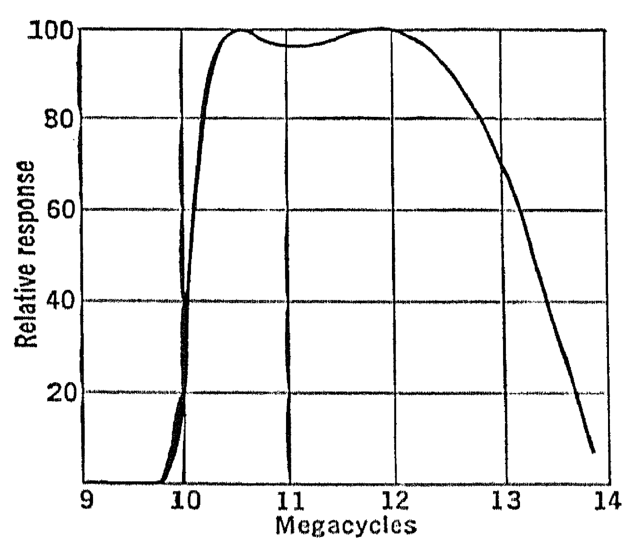


FIG. 17.26.—Overall Response of Intermediate-Frequency Amplifier of an RCA Experimental Television Receiver.

example, may consist of two ordinary stages and two rejector stages. Such a combination gives the approximate response indicated in Fig. 17.26.

There is a definite advantage in employing an automatic gain control in connection with a television receiver. The control voltage for this is obtained by means of a rectifier inserted at the end of the intermediate amplifier, as will be explained in a succeeding section. The voltage from this rectifier regulates the bias on several of the intermediate stages.

The audio intermediate amplifier is usually of a more or less conventional design. It may, for example, consist of two or three stages coupled as shown in Fig. 17.27. Because of the narrower pass band required by the sound, the gain per stage can be made greater in the audio than in the video amplifier. The selectivity needed in the sound intermediate amplifier is governed by the stability of the local oscillator and circuit constancy, rather than by the actual audio range. Unless the pass band is about 100 kilocycles wide, the drift of the local oscillator may produce serious detuning. On the other hand, too broad a band will make tuning

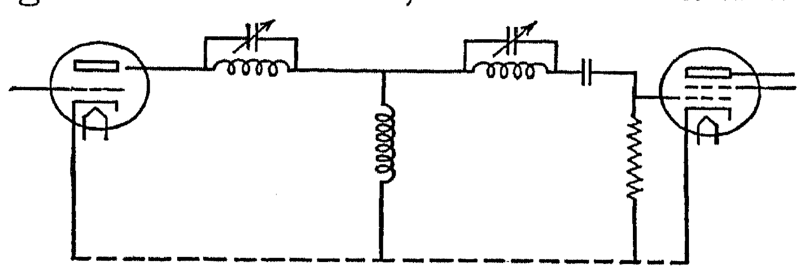


FIG. 17.27.—Audio Intermediate Amplifier Coupling.

of the receiver difficult because the sound is usually used as a basis tuning. The audio second detector and succeeding audio amplifier, loudspeaker, and automatic volume control are no different from those of a broadcast receiver.

The circuits used in experimental receivers constructed in connection with an RCA television field test are given in Fig. 17.28. This diagram includes both the video and audio intermediate amplifiers.

17.6. Detector and Video Amplifier. The signal level from the intermediate amplifier is in the neighborhood of 2 to 5 or more volts. At this level there are a number of types of detectors which are practical. A simple triode plate rectifier or a diode, for example, may be used, or a more complicated full-wave rectifier with its advantage of a higher output resistance may supply the input of the video amplifier. Two representative detector circuits, one with a half-wave rectifier, the other employing a full-wave detector, are illustrated in Fig. 17.29. Just as in a video amplifier stage it is necessary to design the coupling in such a way as to accentuate the upper frequencies of the band to compensate for the capacity of the rectifier. Furthermore, a series inductance in the coupling circuits prevents the intermediate frequencies from entering the first video stage.

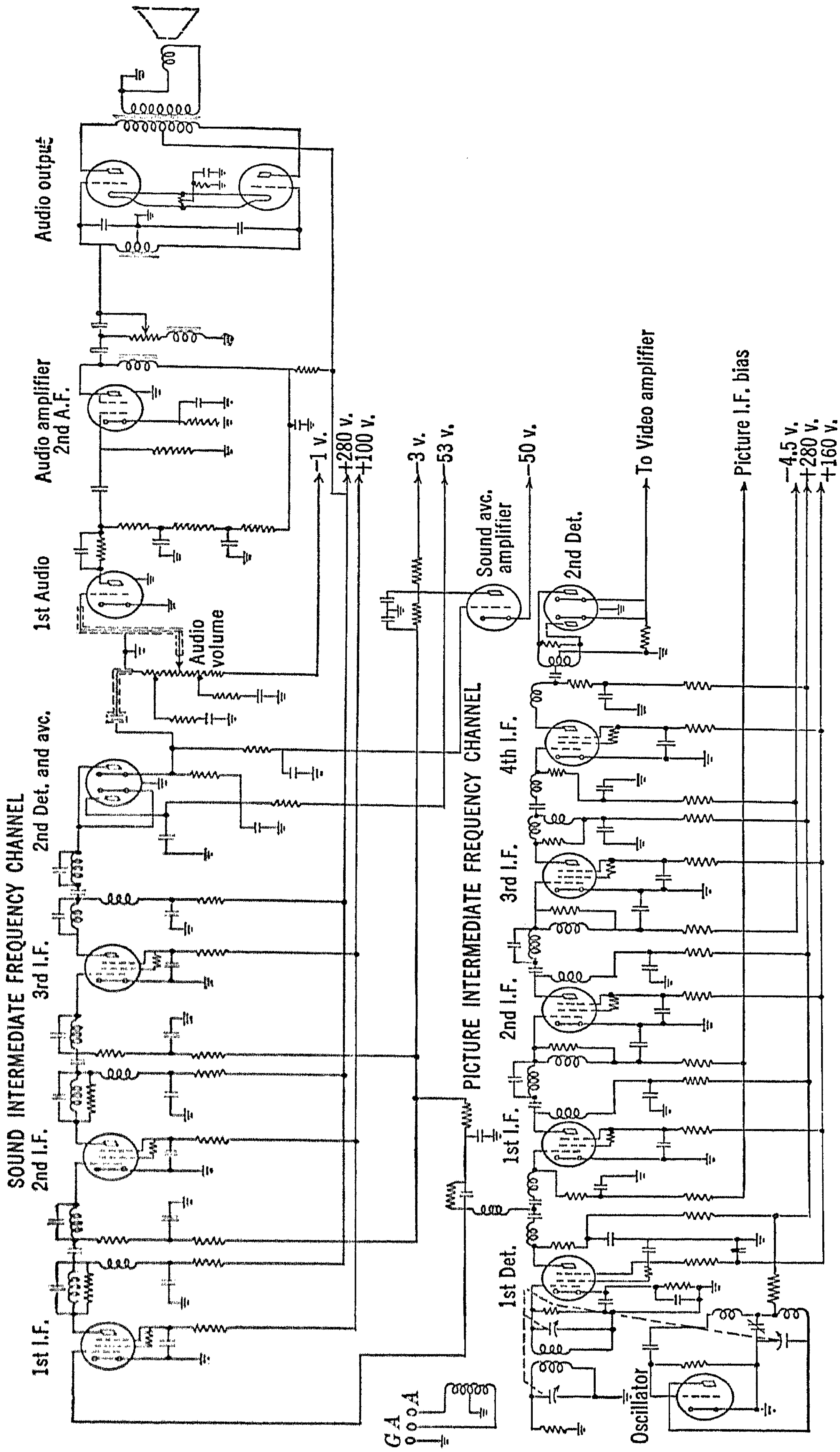


Fig. 17.28.—Radio- and Intermediate-Frequency Circuits of an RCA Experimental Television Receiver.

The polarity of the signal from the second detector depends upon the way in which the rectifier is connected to the intermediate amplifier. If the anode of the rectifier is coupled to the plate of the output tube of the intermediate amplifier the video signal is negative. Reversing the cathode and anode results in a positive signal. In order to produce a positive picture on the screen of the Kinescope an odd number of amplifier stages must follow a detector whose output signal is negative. Similarly an even number of stages must follow if the video signal from the detector is positive. The connections shown in Fig. 17.29 both lead to a negative picture signal.

The video amplifier which follows the second detector makes use of the principles outlined in Chapter 14. Where three stages are found expedient the gain required from each stage is relatively low and uniform response can be obtained simply by using a small plate coupling resistor having a series peaking inductance. An amplifier response which covers a wider

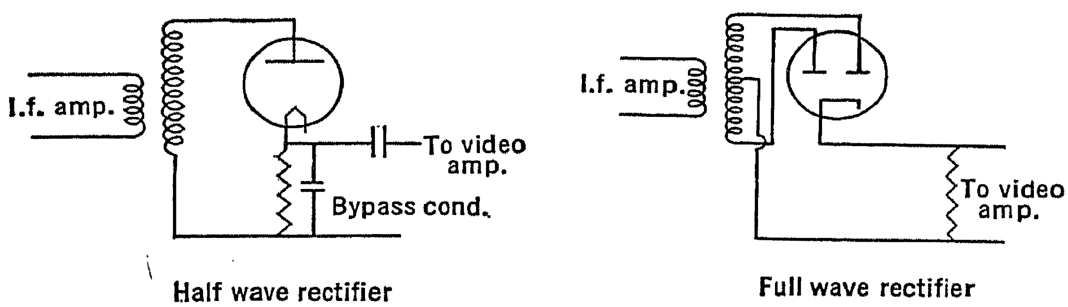


FIG. 17.29.—Typical Second Detector Circuits.

frequency band than the actual video channel does relatively little harm at this point because the signal is well above the noise level, and interfering signals have been eliminated by the selectivity of the intermediate amplifier. The output of this video amplifier actuates the grid of the Kinescope. A peak signal of 20 to 100 volts, depending upon the type of viewing tube, is required for this purpose, working into a load of a few hundred ohms.

Since the video amplifier is essentially an a-c amplifier, it is necessary to reinsert the d-c component of the picture signal at the Kinescope. This is done by making use of the fact that the synchronizing impulse has a constant amplitude above the pedestal and that the pedestal in turn represents black in the picture. The d-c component is usually reinserted at the grid of the tube preceding the Kinescope, this tube being directly coupled to the control grid of the latter.

The automatic gain control voltage is obtained from the intermediate picture signal. This problem is very similar to that of maintaining the d-c level at the Kinescope, and therefore the methods and circuits used

in performing these two functions will be considered together in the next section.

Finally the synchronizing signal must be selected from the picture signal and again divided into horizontal and vertical synchronization. This selection, together with the deflecting circuits themselves, will be considered briefly in a later section.

The terminal tube of the video chain, namely, the Kinescope or viewing tube, needs no further comment. Its construction, the principle of its operation, and the way in which the visible image is formed on its fluorescent viewing screen have already been discussed in detail in Chapter 12.

17.7. The Automatic Gain Control and D-C Level. The automatic volume control of the ordinary broadcast receiver uses the d-c component

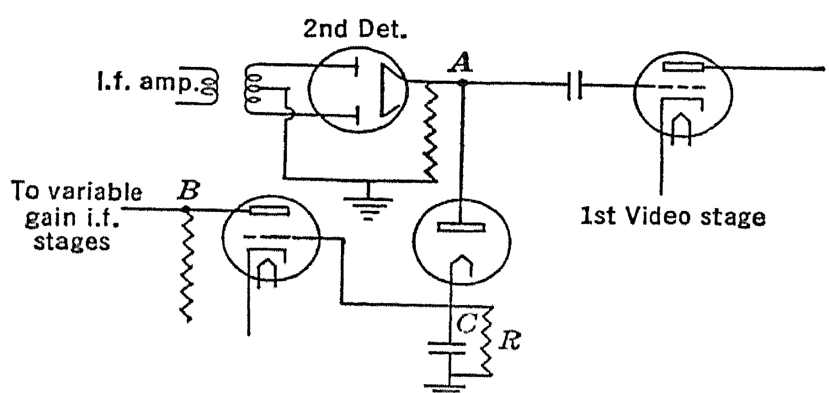


FIG. 17.30.—Automatic Gain Control Circuit.

of the output of the second detector to control the gain of the preceding amplifier. This maintains the average carrier level constant, which is what is required, as the sound signal modulates the carrier about an average value.

Where the d-c component is transmitted by the television transmitter, this procedure cannot be used because the average carrier does not remain constant at the transmitter. Instead, the transmitter maintains the synchronizing impulses at constant amplitude. These, therefore, rather than the carrier, must serve as the reference level for the automatic gain control. The circuit accomplishing this is essentially a peak-reading vacuum-tube voltmeter, followed by sufficient d-c amplification to give the voltage needed to control the variable gain intermediate or r-f amplifier stages. A circuit for carrying this out is shown in Fig. 17.30. The diode rectifier has its anode connected to the output of the second detector, while its cathode is connected to a condenser and resistor in parallel. The time constant of this circuit is made long enough so that

the condenser is not discharged appreciably in a line period. Actually a time constant of the order of 0.01 second is used for the combination of condenser C and resistance R. Point A of the circuit, connected to the second detector, has its maximum excursion in the positive direction at the synchronizing impulse. The cathode of the diode follows its anode as it goes positive, and reaches approximately this maximum value. On the negative swing, the cathode does not follow, except for the small amount by which condenser C is discharged through the resistor R. Therefore the potential across C is approximately the amplitude of the intermediate-frequency signal at the synchronizing impulse. An increase in this amplitude causes an increase in the voltage across C, which, in turn, causes a decrease in the voltage at point B. Point B supplies the bias of the variable gain stages, so that a decrease in its potential reduces the gain of the intermediate- or radio-frequency amplifier.

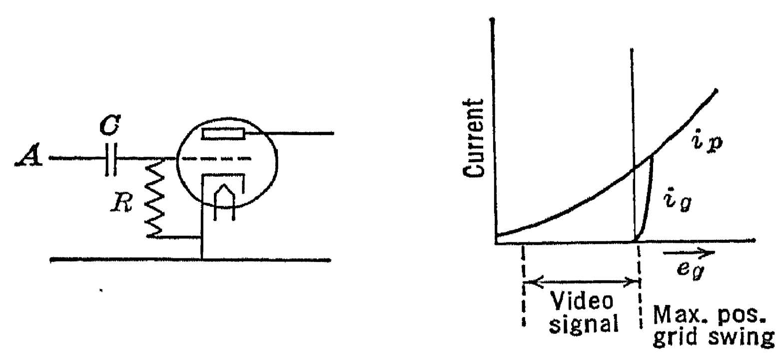


FIG. 17.31.—Restoration of D-C Component of Video Signal.

The reinsertion of the d-c component is based upon a very similar principle. A rectifier establishes the maximum positive swing of the output of the tube preceding the final stage of the video amplifier. This voltage is used to fix the bias of the final tube of the amplifier. In actual practice it is not necessary to use a separate rectifier tube for this purpose, since the grid current in the output tube, if it is run at zero bias, will serve the same end. Fig. 17.31 illustrates the elements of the circuit to carry out the reinsertion, together with a diagram of the action of the circuit. When the signal carries point A to a positive value, the maximum value it can have is limited by the grid current flowing through the large-valued resistor R. The time constant of the RC combination is made such that it does not allow the condenser to discharge appreciably during a line period. Therefore the value of the bias on the output tube will have an absolute value for the peak of each synchronizing impulse, irrespective of the a-c average of the video signal.

17.8. Synchronizing and Deflection. The principles of the deflecting and synchronizing circuits have already been considered in Chapter 15.

Therefore the discussion in this section will chiefly be concerned with these elements in relation to the receiver as a whole.

Before the synchronizing impulses can control the deflection generators they must be separated from the picture signal, and also horizontal and vertical signals must be separated from one another. The synchronizing impulses are obtained from the complete video signal by using a limiter tube, so arranged that only the peaks, as represented by the synchronizing impulse, are passed. The separation of the two types of impulses is carried out by suitable filter circuits, a differentiating net to pass the horizontal, and an integrating net for the vertical.

A fairly high signal level is required to operate most forms of limiter. If, as is true of some receivers, the separating circuits begin at the second detector, several stages of synchronizing amplification may be necessary.

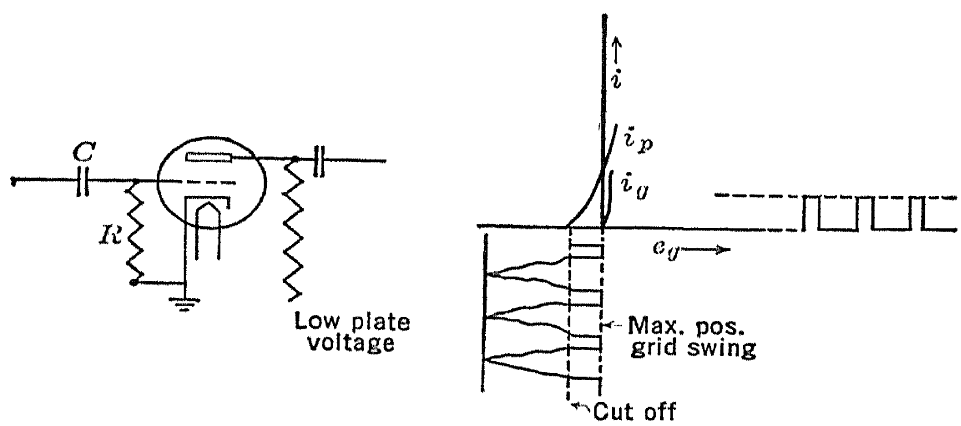


FIG. 17.32.—Limiter for Separating Synchronizing from Video Signal.

However, if the separation is carried out farther up in the video chain no amplification will be needed.

Though a number of limiters might be used, the one diagrammed in Fig. 17.32 is perhaps the simplest and has been found quite satisfactory in test sets. The tube is operated at a low plate voltage so that cutoff occurs when the grid is only slightly negative. The grid is coupled to the synchronizing amplifier through the condenser C and returned to the cathode through the large resistor R. The time constant of this RC circuit is made fairly large so that the most positive peak of the synchronizing impulses brings the grid to zero volts, that is, to the point where the grid starts drawing current. The amplitude of the synchronizing impulses is such that they swing the grid to cutoff. The video signal, which is more negative than the synchronizing signal, is beyond the cutoff of the tube and therefore is not reproduced in the plate current of the separator tube.

The separation of the vertical and horizontal impulses takes place after the video component has been eliminated. Use is made of the difference in duration of the two types of impulses in order to effect the sepa-

ration. The form of the impulses adopted by the R.M.A. for the United States has already been given.

One form of horizontal separator circuit is shown in Fig. 17.33. It is, in fact, a form of high-pass filter, operated over the range of its response where it has a differentiating action on the pulses passing through the circuit. The input and output wave shapes are illustrated in Fig. 17.33*b*. The figure includes the response to the serrated vertical impulse, as well as to the ordinary train of horizontal pulses. It will be noted that the double-frequency equalizing impulses preceding and following the main vertical signal, and the double frequency of the indentations in the main pulse, produce after filtering positive peaks occurring at twice line repetition rate. Those impulses which occur at the midpoint of each line have

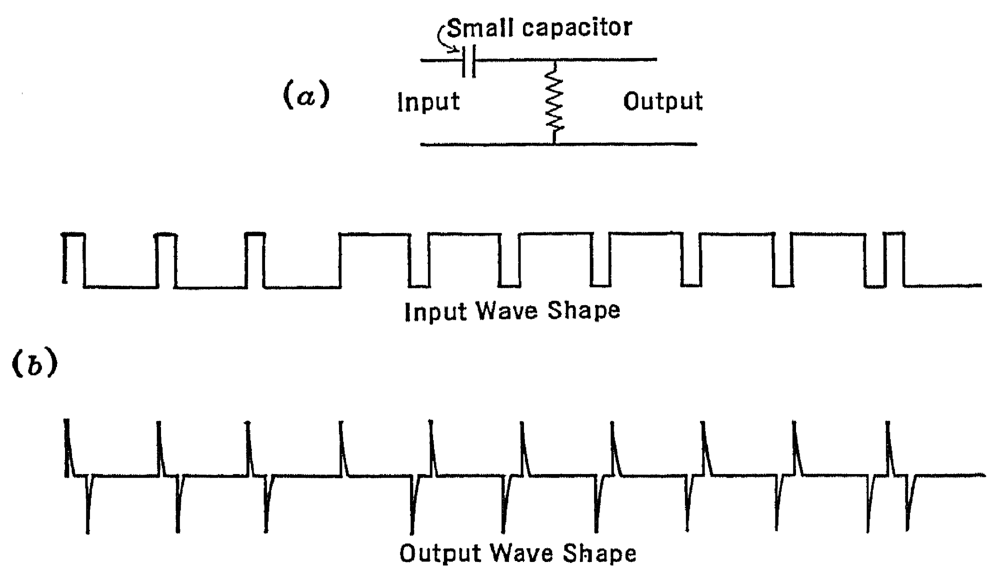


FIG. 17.33.—Circuit for Isolating Horizontal Synchronizing Impulse.

little effect on the horizontal deflection generator since they coincide with a relatively stable part of its cycle. Needless to say, only the positive half of the separator output is used for synchronizing.

The vertical separator is a low-pass filter, used over that part of its range where it has an integrating action. The circuit for this separation is shown in Fig. 17.34, together with the wave shapes for even and odd fields. Equalizing impulses at twice horizontal line frequency are part of the vertical synchronizing impulse. The purpose of the equalizing signals, as has already been pointed out, is to make the synchronizing impulses corresponding to the odd and even field traversals of the interlaced scanning identical. This precaution, together with those necessary to prevent any crosstalk between horizontal pulses from the deflection generator and the input to the vertical selector circuit, must be taken if interlacing is to be maintained.

A typical synchronizing and deflecting circuit is shown in Fig. 17.35.

The horizontal and vertical selector circuits are designated as A and B, respectively. Tube T_A serves as horizontal synchronizing impulse am-

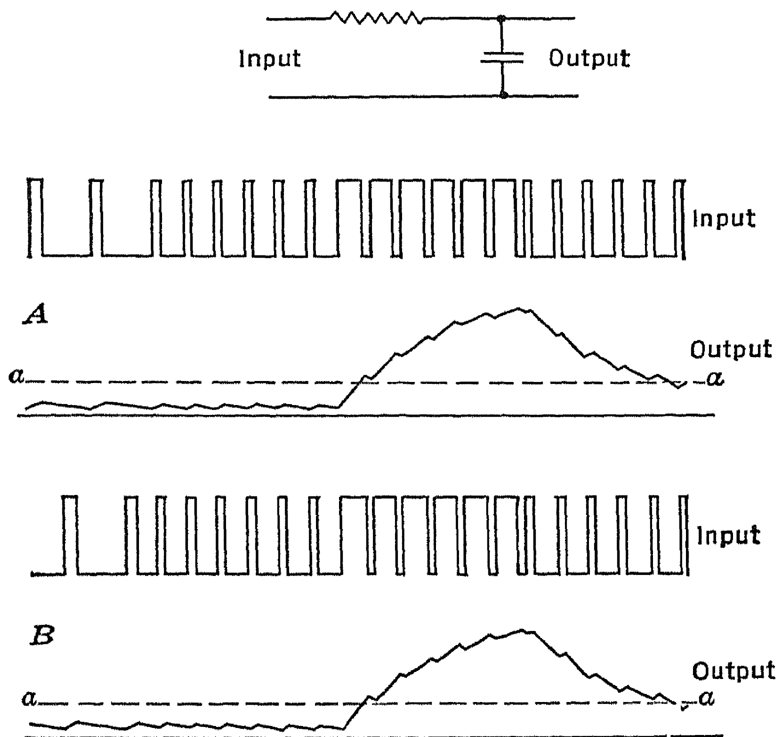


FIG. 17.34.—Selection of Vertical Synchronizing Signal.

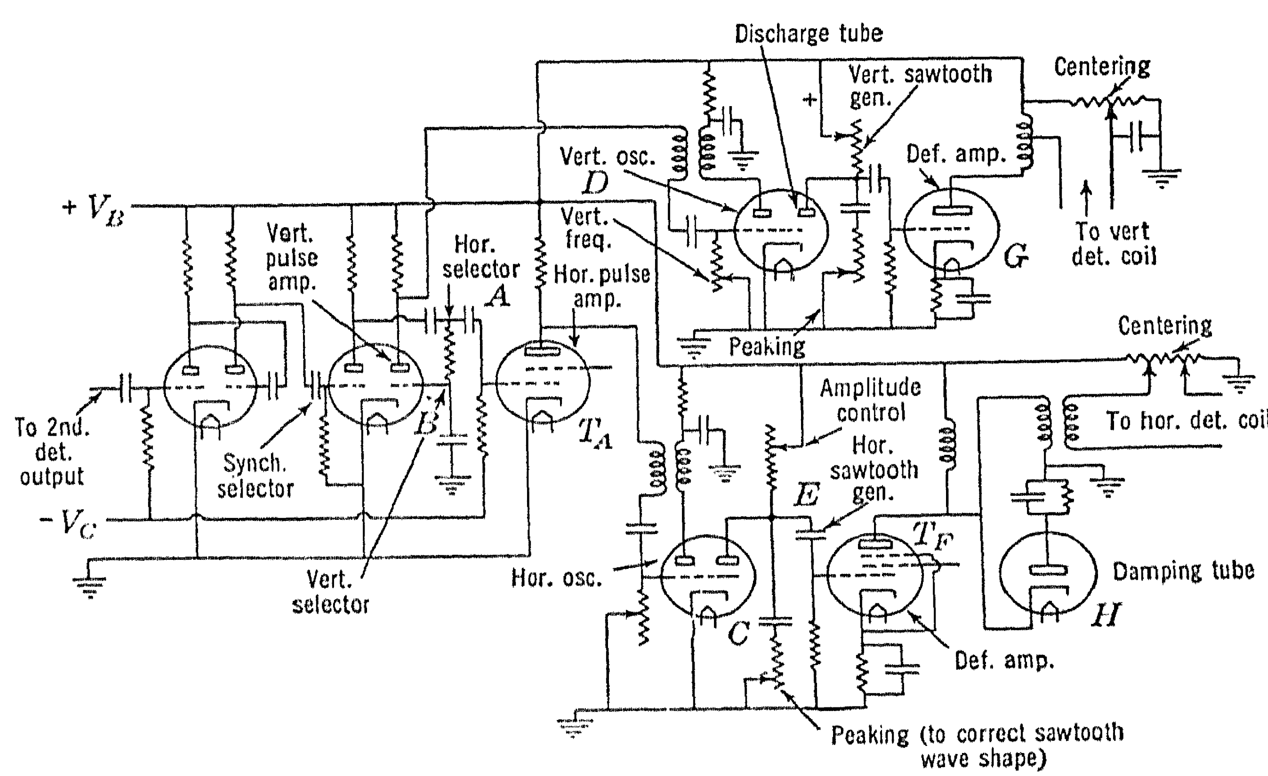


FIG. 17.35.—Typical Synchronizing and Deflecting Circuits.

plifier and as buffer to prevent interaction between the horizontal deflection circuits and the separator filters.

The deflection generators for both vertical and horizontal components of the scanning pattern consist of a relaxation oscillator, a sawtooth generator and discharge tube, and a deflection amplifier.

Blocking oscillators of the type needed to produce the pulses to actuate the discharge tubes are indicated as C and D in Fig. 17.35. The type of oscillator used is a feedback oscillator, arranged with a grid leak and condenser so that its grid is driven to cutoff after one cycle of operation. The grid remains below cutoff until the charge resulting from grid rectification leaks off through the grid leak resistor, permitting another cycle of normal oscillation. The mechanism of operation has been given in a preceding chapter. The frequency of repetition of the cycle is largely determined by the value of the condenser and grid leak. The width of the positive peaks which are used to actuate the discharge tube is determined by the periodicity of the oscillator as a feedback oscillator. In order to control this type of oscillator with the synchronizing signal, its free oscillation frequency must be slightly below the frequency of the synchronizing impulses.

The sawtooth wave shape is formed by charging a condenser through a large resistance and discharging it through a tube whose grid is driven positive by the pulse from the blocking oscillator. In order to make the rise of the potential across the condenser linear, the maximum voltage to which it is charged must be small (in the order of 10 per cent) compared with the voltage applied to the charging resistor. The sawtooth wave is applied either to deflecting plates in the viewing tube, or to the coils of a magnetic deflecting yoke. In all but the least expensive commercial receivers it has been found advantageous to use magnetic deflection for both horizontal and vertical displacement of the scanning spot. Where this type of deflection is employed the current through the coils must have a sawtooth wave shape. In Fig. 17.35 the current for the horizontal deflecting coils is shown supplied from the high-resistance pentode tube T_F, which is used to amplify the output of the sawtooth generator E. Since the inductive load of the coil will have little effect on the current passed by the tube a sawtooth voltage is supplied to its grid. On the other hand, if the amplifier tube is a triode, as in the vertical circuit G, its plate resistance is comparable with the impedance of the coil, and the voltage on its grid should be a sawtooth plus a square wave. This wave shape can be obtained from the sawtooth generator by placing a resistance in series with the charging condenser. A variable resistance is usually inserted at this point, permitting the circuit to be adjusted to give the desired uniformity of deflection.

In order to minimize crosstalk between the horizontal and vertical coils, which would inevitably occur if both windings had a high imped-

ance, due to unavoidable dissymmetry in the yoke construction, one or both of the coils are made to have a relatively low impedance.

When a high-impedance tube supplies a magnetic deflecting unit, it presents very little damping to oscillations that might occur as a consequence of sudden changes in current through the coils. A simple RC damping net cannot be used because it would prevent the building up of the inverse voltage needed to cause the rapid change in current during the return time. In order to obtain the required damping without lengthening the return time, a diode in series with an RC circuit is placed in parallel with the deflecting coils or the driving transformer. This is indicated as H of the horizontal deflecting unit in Fig. 17.35. The vertical deflecting unit, driven by a low plate impedance triode, does not require additional damping across the coils. A resistance in series with the sawtooth generator, however, is needed.

The problems involved in the deflecting yoke itself have already been considered in Chapter 15 and need not be discussed at this point.

17.9. Receiver Voltage Supplies. Two principal voltage sources are needed in the television receiver. One of these supplies the Kinescope voltages; the other, the power required by the amplifiers, deflecting units, etc.

The high-voltage supply is divided into two parts, one which supplies the second anode and screen grid or accelerating electrode voltages, the other the first anode voltage. The former must be capable of delivering about ½ to 1 milliampere at 6000 to 8000 volts. While filtering is required in this rectifier, the percentage ripple which can be tolerated is much higher than is allowable in the rectifier supplying the amplifier. Usually one filter stage consisting, for example, of two 0.03-microfarad condensers and a 30-henry inductance, is ample when used with a half-wave rectifier. The other part of the Kinescope power supply should have a voltage output of 1200 to 1500 volts. This voltage must be variable to permit adjustment of the Kinescope focus. The filtering problem here is identical with that of the second anode voltage supply. The negative ends of both these power units are low so that the output of the video amplifier supplying the control grid can be near ground potential.

The voltage source for the amplifiers resembles those used for sound work. Its output must be much more thoroughly filtered than that of the preceding units or 60-cycle interference will be visible in the reproduced picture. For this purpose usually two or more stages of filtering are required, together with a full-wave rectifier.

17.10. The Complete Receiver. The wiring diagram of a typical home television receiver is shown in Fig. 17.36. In spite of the fact that this diagram has been simplified by omitting some of the incidental detail

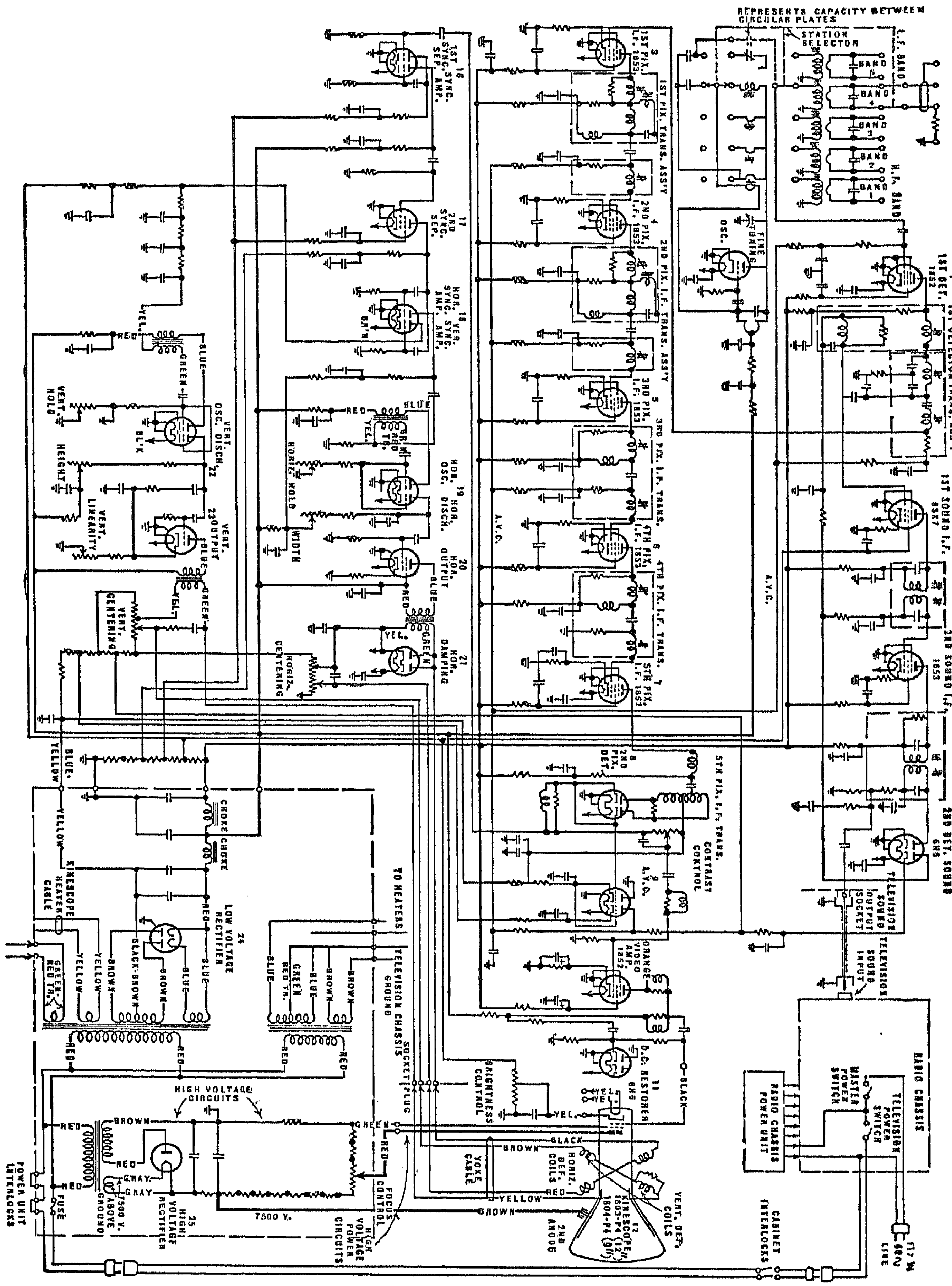


FIG. 17.36.—Circuit Diagram of an RCA Home Television Receiver.

where it is not necessary to illustrate the principles involved, the circuit remains very complex.

The input of the receiver is supplied directly to the first detector without an intervening radio-frequency stage. It will be noticed that station selection is effected by switching to the appropriate pretuned circuit. The output of the detector is separated into audio and video signals. Five video intermediate stages are used, these being controlled by the *AGC*. A full-wave rectifier serves as second detector supplying a single video amplifier stage. The d-c component is reinserted after this stage. It should be noticed that the bias on the Kinescope is obtained through the d-c reinsertion diode so that it is maintained at a negative potential until the cathode of this tube has heated to the point where it gives emission. This is to protect the Kinescope when the set is first turned on, by preventing the beam from reaching the screen before the deflection has started.

The synchronizing signal is taken from the output of the second detector, separated from the video signal, and amplified before horizontal and vertical separation. The deflection generators for the horizontal and vertical scanning are similar in design, except that the former uses a beam power output stage and requires a damping tube, while the latter uses a triode and consequently has a peaking resistor in the sawtooth generator.

17.11. An Experimental Receiver. The construction of an experimental television receiver is not an easy task. Even an experienced worker who is familiar with the building of ultra-high-frequency and broadcast receivers will find his resources severely taxed, because of the complexity of the circuits involved and the accuracy with which they must be constructed. However, it is by no means an impossible undertaking, and a number of successful amateur receivers have been built.

A detailed account of the construction of a television receiver has been published in *QST* by J. B. Sherman* and C. C. Shumard,* describing a very successful outfit which they have assembled and operated. This receiver is representative of the type suitable for "home" construction, and as such it is of no little interest. Owing to limitations of space it is not possible to reproduce their description in full, but sufficient detail can be given to acquaint the reader with the nature of the problem and the equipment required. Those interested in carrying out the actual construction are referred to the original articles. A few changes incorporated in the description below have been suggested by the authors of this series of papers.

* Research Engineering Division RCA Radiotron. See references 24–28.

The original account offers circuits for a number of viewing tubes, including the following tube types:

	Type	Screen Size Inches	Second Anode Voltage Volts	Resolution Lines
RCA	913	1	500	~100
	902	2	600	150
	906	3	1500	250
	1801	5	4000	441
	1802—P2 or P4	5	2000	441
	1804—P4	9	6000	441

The ultra-high, intermediate, and video amplifiers are the same for all these tubes. However, different deflecting circuits and voltage sources are needed.

The circuits needed for the RCA 906 tube will be the only ones discussed in this section, as this is the smallest and least expensive tube from which practical television reception can be had. The very limited resolution of the two smaller tubes gives the pictures reproduced by them little entertainment value.

The power supply for the RCA 906 is shown in Fig. 17.37. Two rectifier tubes are arranged as a voltage doubler so that an ordinary 750-volt receiver power transformer can be used. The condensers needed for the filter can be built up of a series arrangement of 450-volt electrolytic condensers. A list of the circuit constants of the elements used in this power supply is given in Table 17.1.

The synchronizing circuits and deflection generator are shown in Fig. 17.38. Separator circuits of more or less conventional design are used. It will be noticed that, instead of the usual relaxation oscillator and "hard" discharge tubes, self-oscillating gas triodes are used. In spite of the instability of these tubes, which have made them inadvisable in commercial television receivers, they are found satisfactory for an experimental set, particularly because of the simplicity of the circuits involved. The sawtooth outputs from these generators are amplified with medium power pentodes having resistive plate loads coupling them to the horizontal and vertical deflecting plates of the viewing tube. A list of the circuit components of the deflection generator is given in Table 17.2.

Three intermediate amplifier stages, one radio frequency, and one video stage make up the complete picture amplifier. These are arranged as shown in Fig. 17.39. A list of the component parts is given in Table 17.3. The radio-frequency input circuit consists of a tuned grid input, a tuned link, and a single-turn primary loop connected to the antenna transmission line. The plate circuit of the radio-frequency tube contains

TABLE 17.1

C_1—0.001-μfd. mica.
C_2, C_3—1 μfd., 200-volt.
C_4—0.5 μfd., 2000-volt.
C_5—0.5 μfd., 1000-volt.
C_6—16 μfd., 800-volt.
C_7, C_8—4 μfd., 800-volt.
C_9—0.05 μfd., 200-volt.
R_1—1.0 megohm, 1-watt.
R_2, R_4—50,000 ohms, ½-watt.
R_3—0.5 megohm, 1-watt.
R_5, R_6—0.2-megohm potentiometer.
R_7, R_8—0.1 megohm, ½-watt.
R_9—0.5 megohm, 1-watt.
R_{10}—0.5-megohm potentiometer.
R_{11}—0.25 megohm, ½-watt.
R_{12}—50,000-ohm potentiometer.
R_{13}—0.5 megohm, ½-watt.
L_1, L_2—5 henrys.

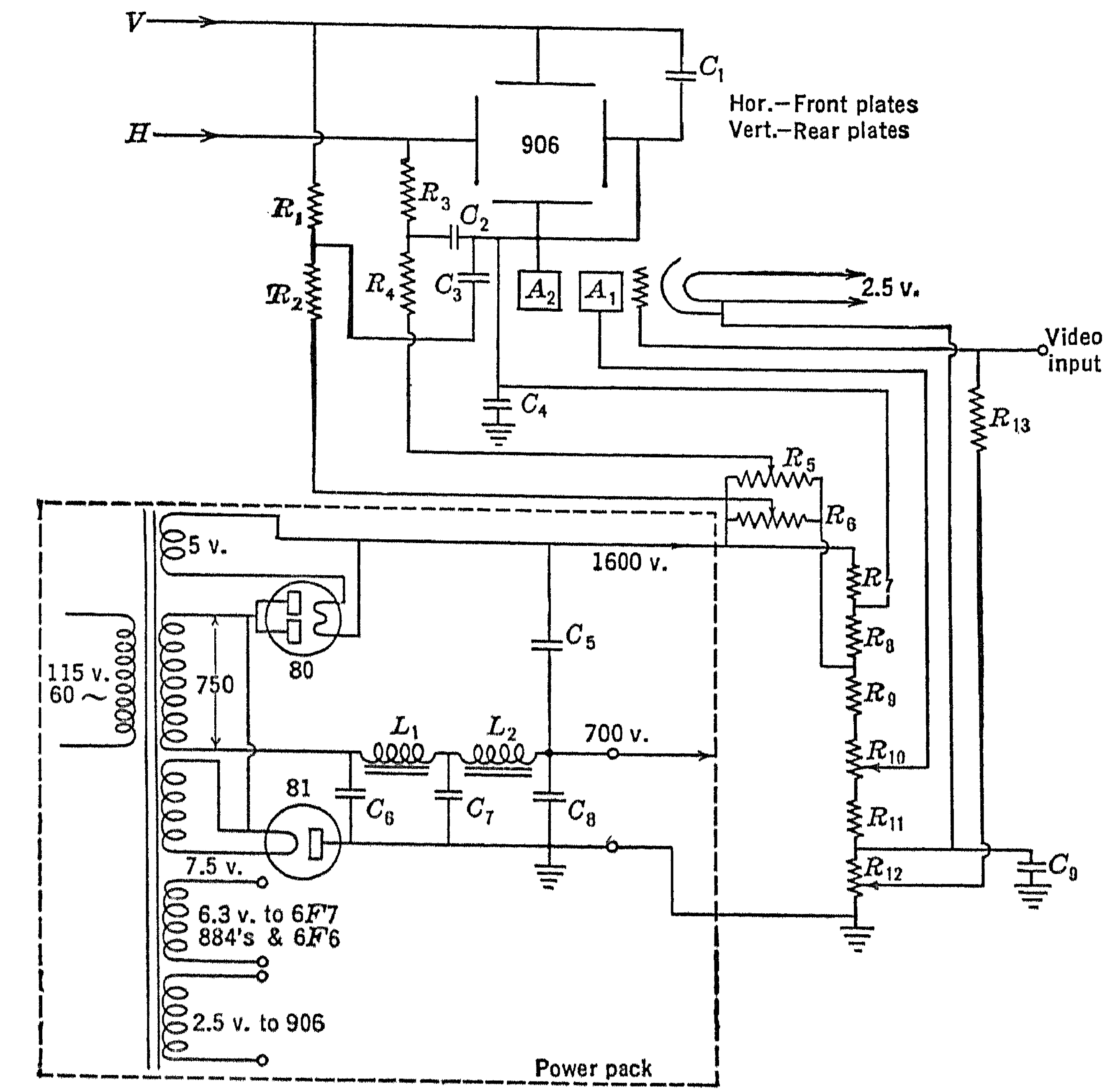


FIG. 17.37.—Power Supply for Experimental Television Receiver.

TABLE 17.2

C_1—0.25 μfd., 400-volt.
C_2—0.001 μfd., 200-volt.
C_3—25 $\mu\mu$fd., 400-volt.
C_4—25 μfd., 25-volt.
C_5, C_7, C_{23}—0.25 μfd., 600-volt.
C_6—0.0005 μfd., 200-volt.
C_8—0.05 μfd., 600-volt.
C_9—50 μfd., 50-volt.
C_{10}—0.5 μfd., 600-volt.
C_{11}—0.05 μfd., 2000-volt.
C_{12}—0.0005 μfd., 2000-volt.
C_{13}—0.01 μfd., 200-volt.
C_{14}, C_{19}—0.1 μfd., 600-volt.
C_{15}—0.002 μfd., 600-volt.
C_{16}, C_{21}—1 μfd., 600-volt.
C_{17}—0.0015 μfd., 600-volt.
C_{18}—50 $\mu\mu$fd., 600-volt.
C_{20}—0.01 μfd., 600-volt.
C_{22}—5 μfd., 50-volt.
R_1, R_6, R_{38}—0.1 megohm, ½-watt.
R_2, R_3, R_{30}—0.2-megohm potentiometer.
R_4—7000 ohms, ½-watt.
R_5—0.2 megohm, ½-watt.
R_7, R_{20}—1 megohm, ½-watt.
R_8, R_{35}—50,000 ohms, ½-watt.
R_9, R_{36}—0.25 megohm, ½-watt.
R_{10}—2000 ohms, ½-watt.
R_{11}, R_{33}—0.15 megohm, 1-watt.
R_{12}, R_{32}—20,000 ohms, ½-watt.
R_{13}—0.4 megohm, ½-watt.
R_{14}, R_{17}—1-megohm potentiometer.
R_{15}—0.1-megohm potentiometer.
R_{16}—2 megohms, ½-watt.
R_{18}, R_{29}—25,000-ohm potentiometer.
R_{19}, R_{24}, R_{25}—0.5 megohm, ½-watt.
R_{21}—0.12 megohm, 2-watt.
R_{22}—25,000 ohms, 10-watt.
R_{23}—0.1 megohm, 4-watt.
R_{26}, R_{27}—10,000-ohm potentiometer.
R_{28}—0.5 megohm potentiometer.
R_{31}, R_{40}—500 ohms, ½-watt.
R_{34}—2500 ohms, ½-watt.
R_{37}—0.15 megohms, ½-watt.
R_{39}—50,000 ohms, 3-watt.

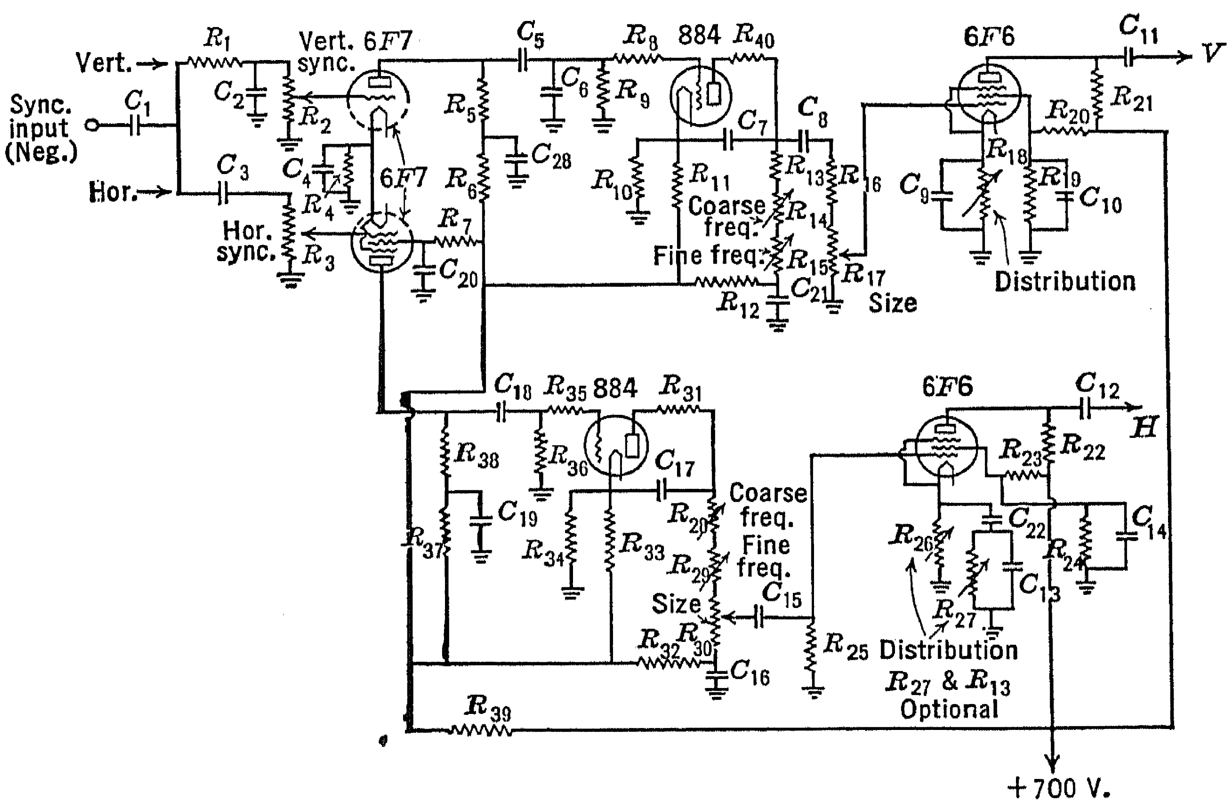


FIG. 17.38.—Synchronizing Circuits and Deflection Generator for Experimental Television Receiver.

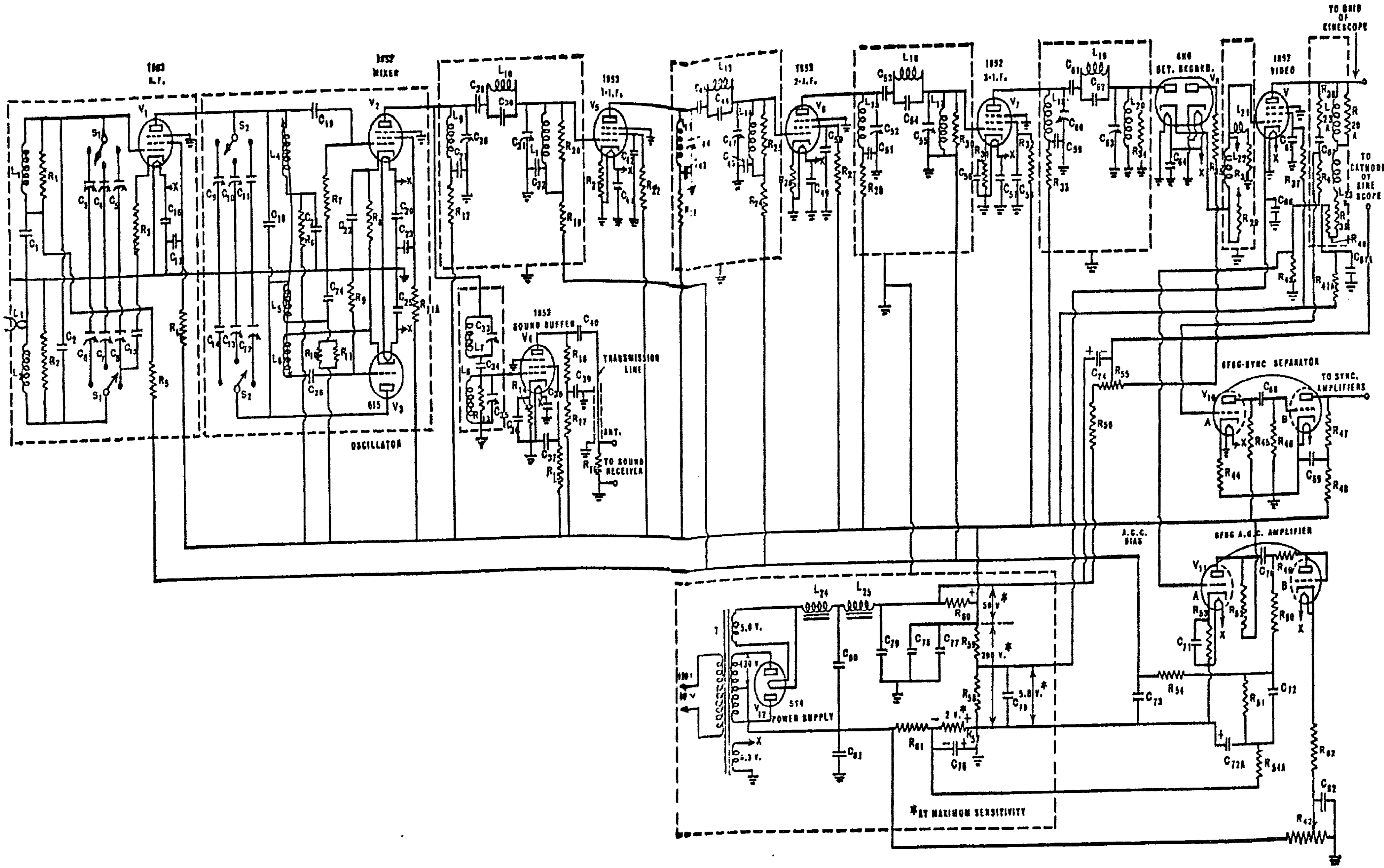


FIG. 17.39.—Complete Amplifier Circuits for Experimental Television Receiver.

TABLE 17.3

C_1, C_{17}, C_{19}, C_{21}, C_{22}, C_{23}, C_{24}—500-$\mu\mu$f mica (Micamold OM900).
C_2—5-$\mu\mu$f mica (Micamold GM800).
C_3, C_4, C_7, C_8, C_9, C_{12}—20-$\mu\mu$f air trimmer (RCA-Victor 12884).
C_5, C_6, C_{10}, C_{13}—12-$\mu\mu$f air trimmer (RCA-Victor 12714).
C_{11}, C_{14}—Value depends on frequency desired.
C_{15}—10-$\mu\mu$f mica (Micamold GM800).
C_{16}, C_{20}, C_{25}, C_{32}, C_{36}, C_{38}, C_{41}, C_{48}, C_{49}, C_{56}, C_{57}, C_{64}—0.01-μf paper, 400-volt (Solar S-0219).
C_{18}—1-$\mu\mu$f (Micamold GM800).
C_{26}—40-$\mu\mu$f mica (Micamold GM800).
C_{27}, C_{37}, C_{39}, C_{40}, C_{42}, C_{43}, C_{50}, C_{51}, C_{58}, C_{59}—0.01-μf, 400-v. paper (Aerovox).
C_{28}, C_{31}, C_{33}, C_{35}, C_{44}, C_{47}, C_{52}, C_{55}, C_{60}, C_{63}—25-$\mu\mu$f variable (Hammarlund APC-25).
C_{29}, C_{34}, C_{45}, C_{53}, C_{61}—560-$\mu\mu$f (RCA-Victor M39, No. 12537).
C_{30}, C_{46}, C_{54}, C_{62}—18-$\mu\mu$f. (See note.) (RCA-Victor M12, No. 12722).
C_{65}, C_{69}—4-μf electrolytic, 450-v. (Aerovox ST-855).
C_{66}—0.01-μf paper + 50-μf elec. 25-v.
C_{67}—0.25-μf (min.) 400-v. paper.
C_{67A}—1-μf 400-v. paper.
C_{68}—0.05-μf, 400-v. paper.
C_{70}—0.004-μf paper.
C_{71}—10-μf electrolytic, 25-v. (Sprague TA-10).
C_{72}—0.05-μf, 200-v. paper.
C_{72A}—50-μf electrolytic, 25-v.
C_{73}—2-μf (min.), 200-v. paper.
C_{74}—8-μf electrolytic, 450-v. (Sprague PTM-8).
C_{75}, C_{76}—50-μf electrolytic, 25-v. (Mallory RS-200).
C_{77}, C_{78}—8-μf electrolytic, 450-v.
C_{79}, C_{80}—16-μf electrolytic, 450-v.
C_{81}, C_{82}—50-μf electrolytic, 25-v.

R_1, R_{34}—3000 ohms, 0.5-watt.
R_2, R_7, R_{17}, R_{39}, R_{40}, R_{41A}—2000 ohms, 0.5-watt.
R_3, R_{21}, R_{26}—68 ohms, 0.5-watt.
R_4, R_{10}, R_{11}, R_{15}, R_{22}, R_{27}, R_{54}—30,000 ohms, 0.5-watt.
R_5, R_{19}, R_{24}, R_{29}, R_{29A}, R_{49}—20,000 ohms, 0.5-watt.
R_6, R_8, R_{12}, R_{18}, R_{23}, R_{28}, R_{33}, R_{44}, R_{53}, R_{54A}—1000 ohms, 0.5-watt.
R_9, R_{47}—10,000 ohms, 0.5-watt.
R_{11A}—150,000 ohms, 0.5-watt.
R_{13}, R_{56}—15,000 ohms, 0.5-watt.
R_{14}, R_{43}—200 ohms, 0.5-watt.
R_{16}—50 ohms, 0.5-watt.
R_{20}, R_{25}, R_{30}—2500 ohms, 0.5-watt.
R_{31}—160 ohms, 0.5-watt.
R_{32}—60,000 ohms, 1-watt.
R_{35}, R_{51}—200,000 ohms, 0.5-watt.
R_{36}—4000 ohms, 0.5-watt.
R_{37}—50,000 ohms, 1-watt.
R_{38}—50,000 ohms, 0.5-watt.
R_{41}—5000 ohms, 0.5-watt.
R_{42}—5000-ohm potentiometer.
R_{45}, R_{52}—20,000 ohms, 1-watt.
R_{46}—3 megohms, 0.5-watt.
R_{48}—1 megohm, 0.5-watt.
R_{50}—50,000 ohms, 0.5-watt.
R_{55}—25,000-ohm potentiometer.
R_{57}—20 ohms, 1-watt.
R_{58}—100 ohms, 1-watt.
R_{59}—5000 ohms, 25-watt.
R_{60}—340 ohms, 10-watt.
R_{61}—50 ohms (or more), 10-watt.
R_{62}—500 ohms, 0.5-watt.
L_1 to L_{23}—described in text.
L_{24}, L_{25}—20-henry, 200-ma, 100-ohm, filter choke.
T_1—Power transformer (Thordarson Type T-13R16).
S_1, S_2—Mallory-Yaxley selector switch No. 3243-J

NOTE: Value of these condensers is critical.
All fixed ½- and 1-watt resistors are I. R. C. Type BT.

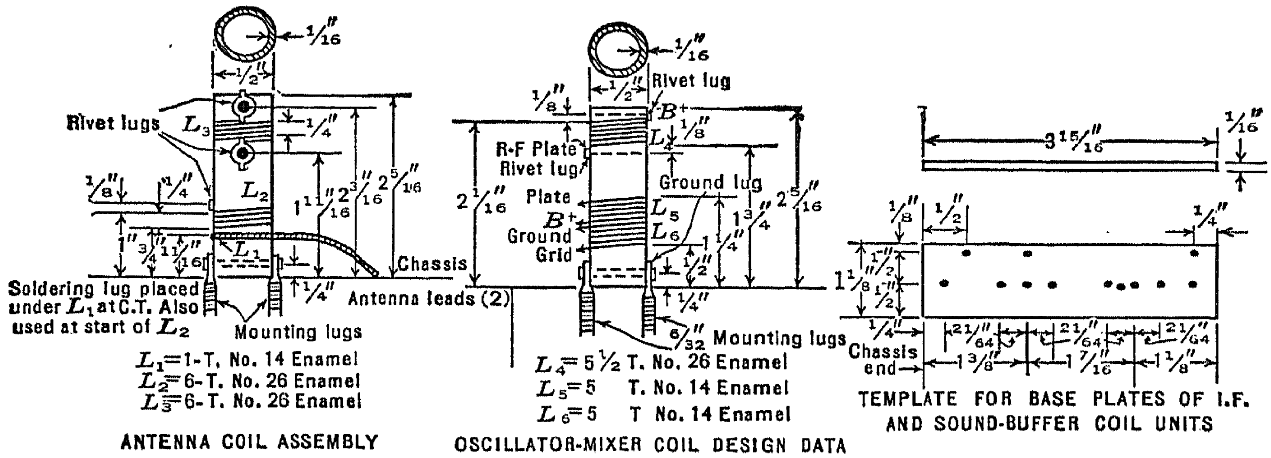


FIG. 17.40.—Coil Design Data I.

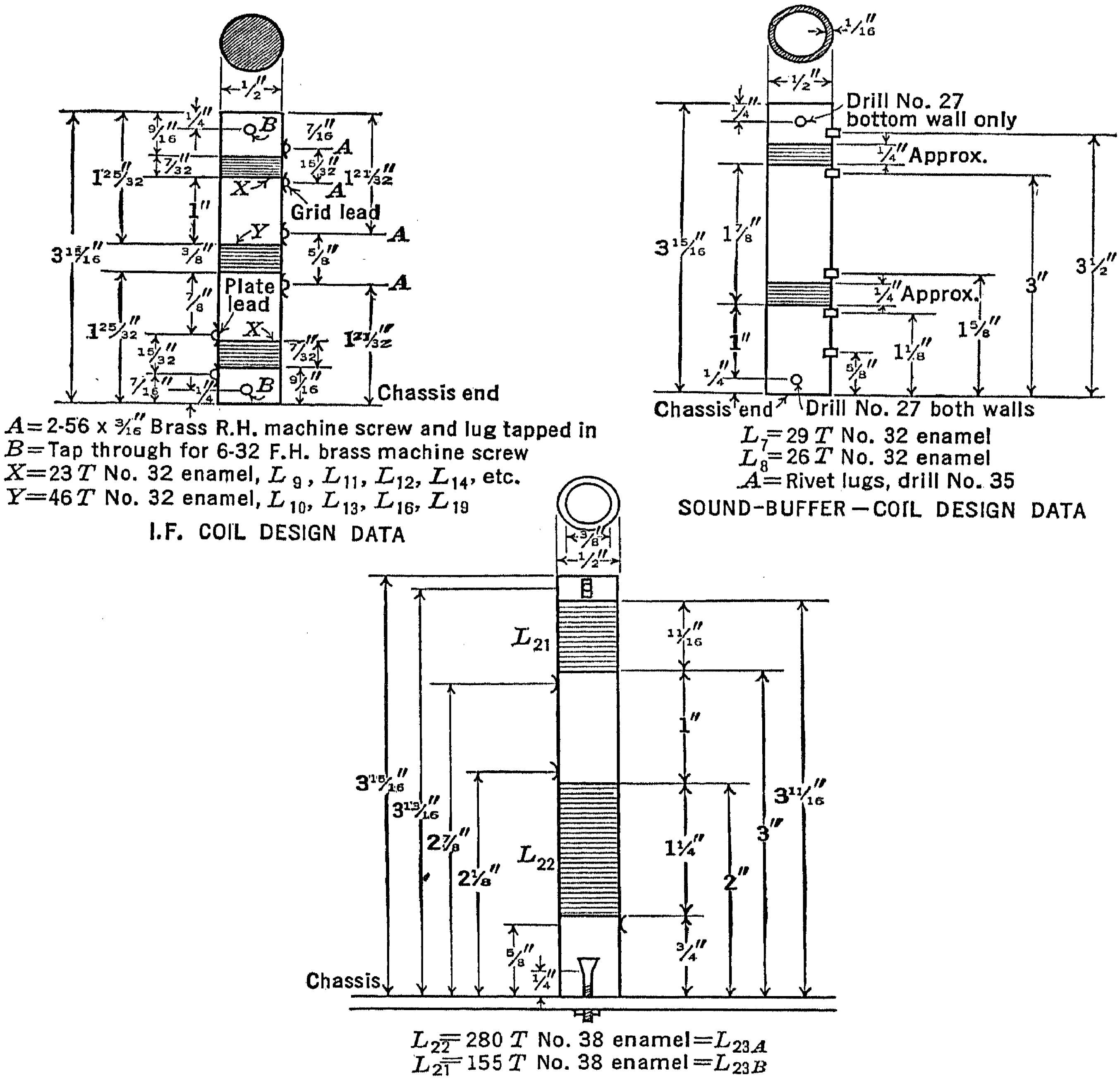


FIG. 17.41.—Coil Design Data II.

a simple tuned circuit, coupling it to the grid of the RCA 1852 which is used as the mixer. The coupling between the RCA 6J5 local oscillator and the mixer is partly capacitative and partly inductive. The intermediate video carrier is adjusted for 13.0 megacycles, and the single sideband response is about 3 megacycles wide. A gain of 5 to 8 per stage can be obtained over this pass band. More or less conventional circuits are used for the second detector and radio amplifier. The synchronizing signal is obtained from the video output, amplified by a single triode stage and then separated from the video signal. The coils needed for the radio- and intermediate-frequency inductances have to be made rather carefully if excessive losses and unwanted feedback are to be avoided. Diagrams of the construction of these coils are shown in Fig. 17.40 and Fig. 17.41.

Alignment of the radio and intermediate frequency is difficult and requires test equipment such as a calibrated oscillator, a radio-frequency vacuum-tube voltmeter, and an ultra-high-frequency receiver. Detailed instructions for carrying out the alignment procedure will be found in the articles referred to.

Although the construction of this receiver is difficult and laborious, it will nevertheless be valuable experience in the practical aspects of the types of circuits involved in television work. The training afforded by carrying out a project of this nature would be of no little value to anyone planning to design or service more elaborate television receivers.

BIBLIOGRAPHY

1. J. C. Wilson, "Television Engineering," Pitman, London, 1937.
2. "Fernsehen," F. Schröter (editor), J. Springer, Berlin, 1937.
3. M. v.Ardenne and D. S. Puckle, "Television Reception," Van Nostrand, New York, 1937.
4. F. E. Terman, "Radio Engineering," McGraw-Hill, New York, 1937.
5. "Television Today," George Newnes Ltd., London, 1936.
6. V. K. Zworykin, "Experimental Television System and Kinescopes," *Proc. I. R. E.*, Vol. 21, pp. 1655–1673, December, 1933.
7. R. S. Holmes, W. L. Carlson and W. H. Tolson, "Experimental Television System—Part III, The Receiver," *Proc. I. R. E.*, Vol. 22, pp. 1266–1285, November, 1934.
8. R. R. Beal, "Equipment Used in the Current RCA Television Field Tests," *R. C. A. Rev.*, Vol. 1, pp. 36–48, January, 1937.
9. E. W. Engstrom and R. S. Holmes, "Television Receivers," *Electronics*, Vol. 11, pp. 28–31, 63–66, April, 1938.
10. E. W. Engstrom and R. S. Holmes, "Television I. F. Amplifiers," *Electronics*, Vol. 11, pp. 20–23, June, 1938.
11. E. W. Engstrom and R. S. Holmes, "Television V. F. Circuits," *Electronics*, Vol. 11, pp. 18–21, August, 1938.
12. E. W. Engstrom and R. S. Holmes, "Television Synchronization," *Electronics*, Vol. 11, pp. 18–20, November, 1938.

13. E. W. Engstrom and R. S. Holmes, "Deflection Circuits in Television Receiver," *Electronics*, Vol. 12, pp. 19–21, 32, January, 1939.
14. E. W. Engstrom and R. S. Holmes, "Power for Television Receivers," *Electronics*, Vol. 12, pp. 22–24, April, 1939.
15. E. W. Engstrom, "Television Receiving and Reproducing Systems," *J. Applied Physics*, Vol. 10, pp. 455–464, July, 1939.
16. S. W. Seeley, "Effect of the Receiving Antenna on Television Reception Fidelity," *R. C. A. Rev.*, Vol. 2, pp. 433–441, April, 1938.
17. H. T. Lyman, "Television R-F Input Circuits," *R. M. A. Engineer*, Vol. 3, pp. 3–6, November, 1938.
18. B. J. Thompson, "Review of Ultra-High-Frequency Vacuum-Tube Problems," *R. C. A. Rev.*, Vol. 3, pp. 146–155, October, 1938.
19. W. R. Ferris, "Input Resistance of Vacuum Tubes as Ultra-High-Frequency Amplifiers," *Proc. I. R. E.*, Vol. 24, pp. 82–107, January, 1936.
20. A. P. Kauzmann, "New Television Amplifier Receiving Tubes," *R. C. A. Rev.*, Vol. 3, pp. 271–289, January, 1939.
21. W. T. Cocking, "Television I. F. Amplifiers," *Wireless Engineer*, Vol. 15, pp. 358–362, July, 1938.
22. D. G. Fink, "A Laboratory Television Receiver—Parts I to VI," *Electronics*, Vol. 11, pp. 16–20, July; 26–29, Aug.; 22–25, Sept.; 16–19, Oct.; 26–29, Nov.; and 17–19, Dec., 1938.
23. "Television Receivers in Production," *Electronics*, Vol. 12, pp. 23–25, March 1939.
24. J. B. Sherman, "Building Television Receivers with Standard Cathode Ray Tubes," *QST*, Vol. 22, pp. 21–25, October, 1938.
25. C. C. Shumard, "A Practical Television Receiver for the Amateur," *QST*, Vol. 22, pp. 21–25, 72–76, December, 1938.
26. C. C. Shumard, "Construction and Alignment of the Television Receiver," *QST*, Vol. 23, pp. 45–52, 116–118, January, 1939.
27. J. B. Sherman, "Using Electromagnetic-Deflection Cathode Ray Tubes in Television Receiver," *QST*, Vol. 23, pp. 40–44, 106, February, 1939.
28. J. B. Sherman, "An Electrostatic-Deflection Kinescope Unit for the Television Receiver," *QST*, Vol. 23, pp. 52–55, March, 1939.
29. "Practical Television," RCA Manufacturing Co., Camden, N. J., 1939.

PART IV

RCA-NBC Television Project

CHAPTER 18

RCA TELEVISION PROJECT—STUDIO AND MONITORING EQUIPMENT

The foregoing chapters have described the various elements which go to make up a television system. These elements, considered in detail individually as they have been, may seem to be essentially unrelated rather than the parts of a single unit. In order not to lose perspective on the coordinated whole, the following description of a working example of a television system is given. Because of the authors' familiarity with this particular installation, the equipment used in the field test which the Radio Corporation of America has undertaken prior to initiating routine television broadcasting is chosen for illustration. These field tests have been undertaken at frequent intervals during the past ten years. The installation described represents the result of very extensive research, culminating in not only the development of the television equipment required for broadcasting, but also the studio and operating technique so necessary for the production of a picture of high entertainment value.

18.1. Preliminary Television Field Tests. After many years of research and development an all-electronic television system emerged from the laboratory in 1933 for actual field tests. These tests were carried out at Camden, using a video transmitter with a carrier frequency of 49 megacycles. The accompanying sound was transmitted at 50 megacycles. The studio from which the picture signals originated was located about 1000 feet from the transmitter and connected to it by a coaxial line. Iconoscopes were used to pick up scenes both in the studio and out-of-doors. A scanning pattern of 240 lines made it possible to obtain a picture with good definition, but as the frame frequency was 24 cycles, without interlacing, flicker was quite noticeable.

The following year the number of lines was increased to 343, and an interlaced pattern having a field frequency of 60 cycles and a repetition rate of 30 frames per second was adopted. The results of these tests were so satisfactory that it was decided to continue them in New York City, the site of earlier RCA tests using a mechanical scanner. The advantage of the new location was that transmission studies under more nearly the conditions encountered in actual broadcasts were possible, in particular,

with respect to noise and reflection from buildings. This move was made in 1935, tests starting in June the following year.

18.2. Tests by the National Broadcasting Company. The New York studios were located in Radio City. The transmitter was installed in one of the upper floors of the Empire State Building, with the antenna on the mooring mast, 1285 feet above street level. Two links interconnect the studio and transmitter. One of these is an underground coaxial cable approximately a mile in length. An ultra-high-frequency radio relay link operating at 177 megacycles serves as alternative for interconnecting the two units. In order to increase the flexibility of the system, and to permit outdoor and indoor pickup from remote points, a mobile unit consisting of a pickup truck and a transmitter, which operated at 177 megacycles, was placed in service in 1938.

Approximately one hundred receivers were built and located at various points within a radius of 50 miles of the transmitter. These, together with field strength measurements, gave detailed information as to the effect of the terrain on the received pictures. They also facilitated obtaining data on the reaction of a great variety of people to different types of programs.

Of course it should be realized that neither the transmitting equipment nor the receivers were constructed into a state of static completeness at their inception. Actually all this equipment has been in a process of evolution ever since it was installed. As the result of improvement in equipment, operating technique and programming, the picture that is being transmitted today is incomparably better than that of 1936, when transmission was started.

The television equipment at Radio City may be conveniently separated into three units. The direct pickup studio is located on the third floor, together with its associated control room. Above it on the fifth floor is the film studio with its projectors, cameras, and control apparatus. Finally, on this same floor there is an equipment room which contains the synchronization and deflection generators controlling the whole system and the line amplifiers.

18.3. Direct Pickup Studio. The direct pickup studio itself is a room 50 by 30 feet and 18 feet in height. The length of the studio is so chosen that full advantage may be taken of the lenses which can conveniently be adapted to an Iconoscope camera. The shortest focal length for lenses ordinarily used with the Iconoscope is approximately 6½ inches, having an angular field of 37°. At 50 feet the field of vision is nearly 30 feet. Thus with the size and shape of studio selected the scene being televised may have any width up to 30 feet.

Although there is no fixed size or arrangement for the studio scenes,

the room is designed to accommodate one large set and at least one additional smaller one. The need for having two or more adjoining sets is obvious when it is remembered that, unlike the making of motion-picture films, where discontinuities between one scene and the next are of no importance, a television performance must be continuous, with an absolute minimum of interruptions. Where two or more sets are available one can be televised while the scenery of the others is being changed. Of course, films are available and may be used to avoid program interruptions, but these may not always be convenient. The two alternatives increase the flexibility and add to the range and variety of play that can be produced.

The studio ceiling is made quite high for several important reasons. It makes for better and more flexible acoustical properties, it permits the use of overhead lighting, thus clearing the floor for more important uses, and finally it facilitates air-conditioning and cooling the room.

The accompanying sound is, of course, very important, and the studio must be constructed with due regard to acoustical principles. Reverberation time is one of the governing factors in obtaining desired sound effects. The volume of the studio and the area and reflectivity of the boundary walls are the controlling factors of the time of reverberation. In order to vary the reverberation so that different effects can be obtained, the effective absorption is made adjustable. The ceiling and end walls of the studio are covered with rockwool over which is laid perforated sheet metal. The side walls, instead of being treated in this way, are equipped with movable absorbing panels. Three pairs of these panels are arranged along each of the side walls, each pair of panels being 10 feet wide and 15 feet high, and covered with the same material used on the end walls and ceiling. The panels, when not wanted, slide behind sheet-metal pilasters, leaving exposed the highly reflecting walls behind them. In order to avoid reverberation from the pilasters themselves, the sheet metal is backed up with cork. Direct sound reflections must also be avoided; therefore the walls back of the panels and the faces of the pilasters are fluted. When the panels cover the side walls of the studio the reverberation time for a frequency of 1000 cycles per second is approximately 0.4 second. By opening the panels the reverberation time is nearly doubled. When the scene being transmitted calls for a greater reverberation than can be obtained by opening the panels, the echo chambers found so successful for sound broadcasting may be used. The panels are hydraulically operated and can be moved by remote control from the control room.

The technique of handling the microphones for television pickup is somewhat more difficult than that for sound broadcasting, because not only must the microphones be kept out of view of the cameras, but also

the actors' positions are fixed by the visible scene so that they cannot move towards or away from them in order to obtain a desired sound effect, nor group themselves to produce the greatest realism. In order to overcome this handicap, the microphones must be more maneuverable than is ordinarily necessary. To attain this end a boom microphone is used. This consists of a velocity microphone, equipped with a windshield, which is suspended from a long arm. The arm, or boom, can be moved over the scene of action, and the microphone raised or lowered as desired.

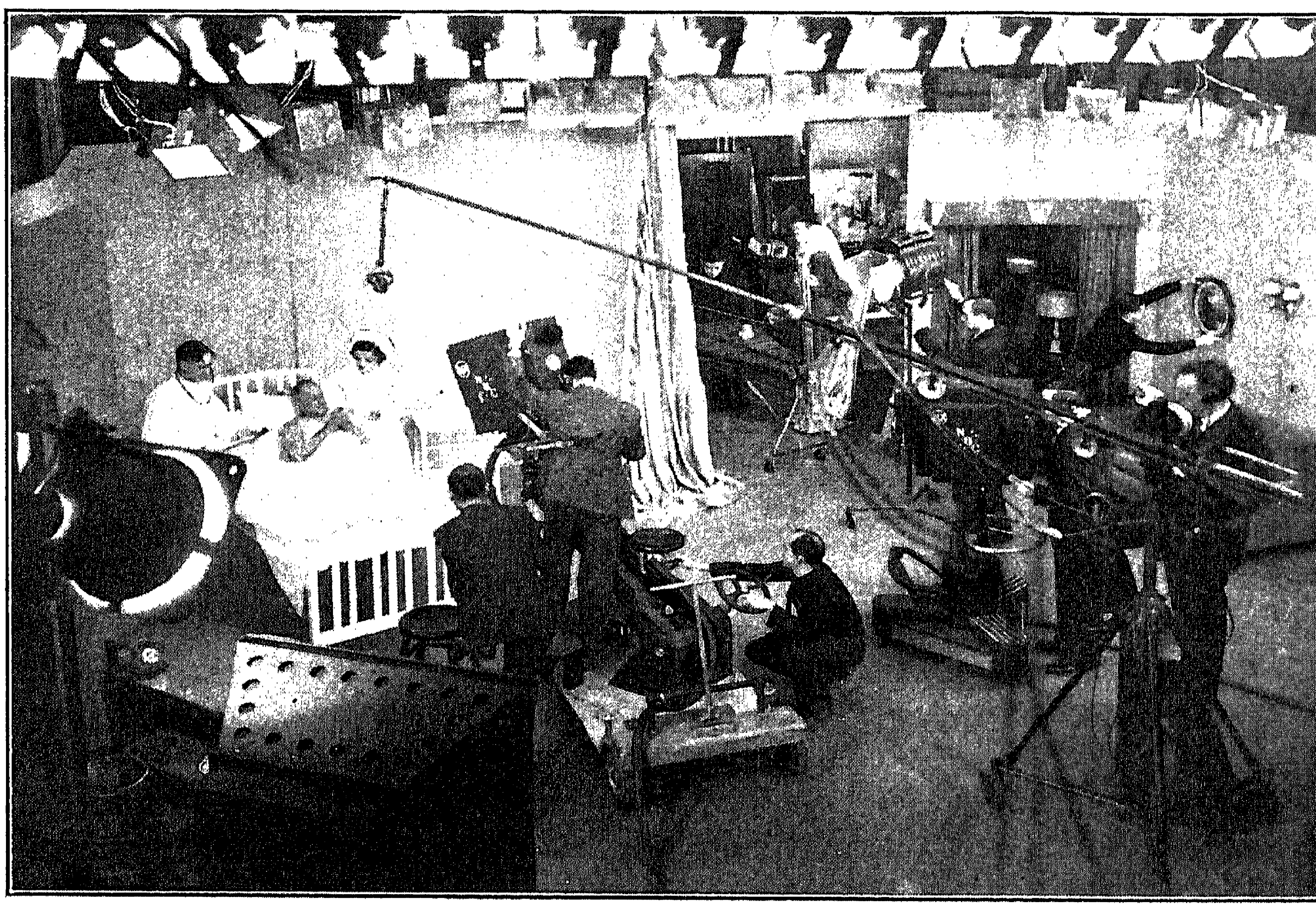


FIG. 18.1.—Studio Scene Showing Use of Boom Microphone (courtesy of NBC).

A boom microphone is clearly in evidence in the studio scene shown in Fig. 18.1. Parabolic reflector microphones are also employed, which, though lacking the maneuverability of the boom microphone, are more easily adjusted and require less attention.

The lighting equipment for this studio was planned to give the greatest flexibility possible, and yet not to interfere with the positioning of the cameras or sets.

In order to obtain an optimum picture and yet permit stopping down the lens sufficiently to obtain a large depth of focus, it has been found expedient to use a fairly high level of illumination, that is to say, 1000 to 2000 foot-candles. This lighting must be distributed in order to avoid

harsh shadows. Therefore, the illumination, instead of being obtained from a few high-power lamps, is obtained from a great many small units. Banks of 300-watt lamps are mounted on metal frames of various sizes and shapes. Each lamp is equipped with a reflector, which is an integral part of the bulb, and is frosted to diffuse the light. The frames carrying the lamps are suspended from the ceiling of the studio by means of cords and pulleys in such a way that they can be not only raised or lowered, but also tilted and moved laterally. Enough of these banks are available to meet any situation that may be encountered, but of course it may not be necessary to use more than a fraction of them at any one time. In addition to the diffuse lighting the studio is equipped with spot-, modeling-, and searchlights of various candlepower, including several special large units, so that shadow effects as desired can be obtained. Incandescent lamps have been selected as the most satisfactory means of producing the illumination. This selection is the result of tests on nearly all the now commercially available high-intensity light sources. The new fluorescent lamps, however, may eventually prove even more satisfactory. Also high-pressure mercury lamps have shown considerable promise. At present the lights are operating on direct current obtained from generators located in the building. The total power consumption is more than 50 kilowatts when all the lights are on, but is usually somewhat less than this.

One of the most important points of studio technique to be investigated during the course of the "field tests" was the question of lighting, and therefore many changes in the systems of illumination have been made as the tests progressed. The present lighting arrangement, described above, can be controlled easily and rapidly, gives ample illumination without undue glare, and, whether or not it is the final form, permits obtaining a very satisfactory picture with a wide variety of sets. The lighting arrangement for a typical studio set is illustrated in Fig. 18.2.

The generation of the light required in the studio is of course accompanied by the development of a large amount of heat. In order to maintain a comfortable working temperature the studio is equipped with an air-conditioning system capable of carrying away this heat. The design of a cooling system capable of handling the heat load without excessive temperature gradients and without too rapid a change of air proved somewhat difficult. However, these difficulties have been overcome in the present system, which has proved very satisfactory, operating as it does with a temperature differential of about 30°F.

Where the set requires the concentrated light from several high-power spots, the infra-red radiation accompanying the visible light may be sufficient to produce serious heating in the set itself. This is avoided by heat filters in front of the lights, which remove most of the heat from the beam

and diffuse it to the room. The heat filters are made of special glass, in the form of strips placed side by side, instead of a single sheet, to minimize the danger of cracking due to stresses produced by heat gradients.

Besides the ordinary movable stage properties, the studio is arranged so that scenic backdrops and ordinary scenery flats can be used. These are hung from a network of pipes crisscrossing the ceiling, which pipes also support the overhead lighting. The end wall of the studio facing the cameras incorporates a projection booth, from which may be projected

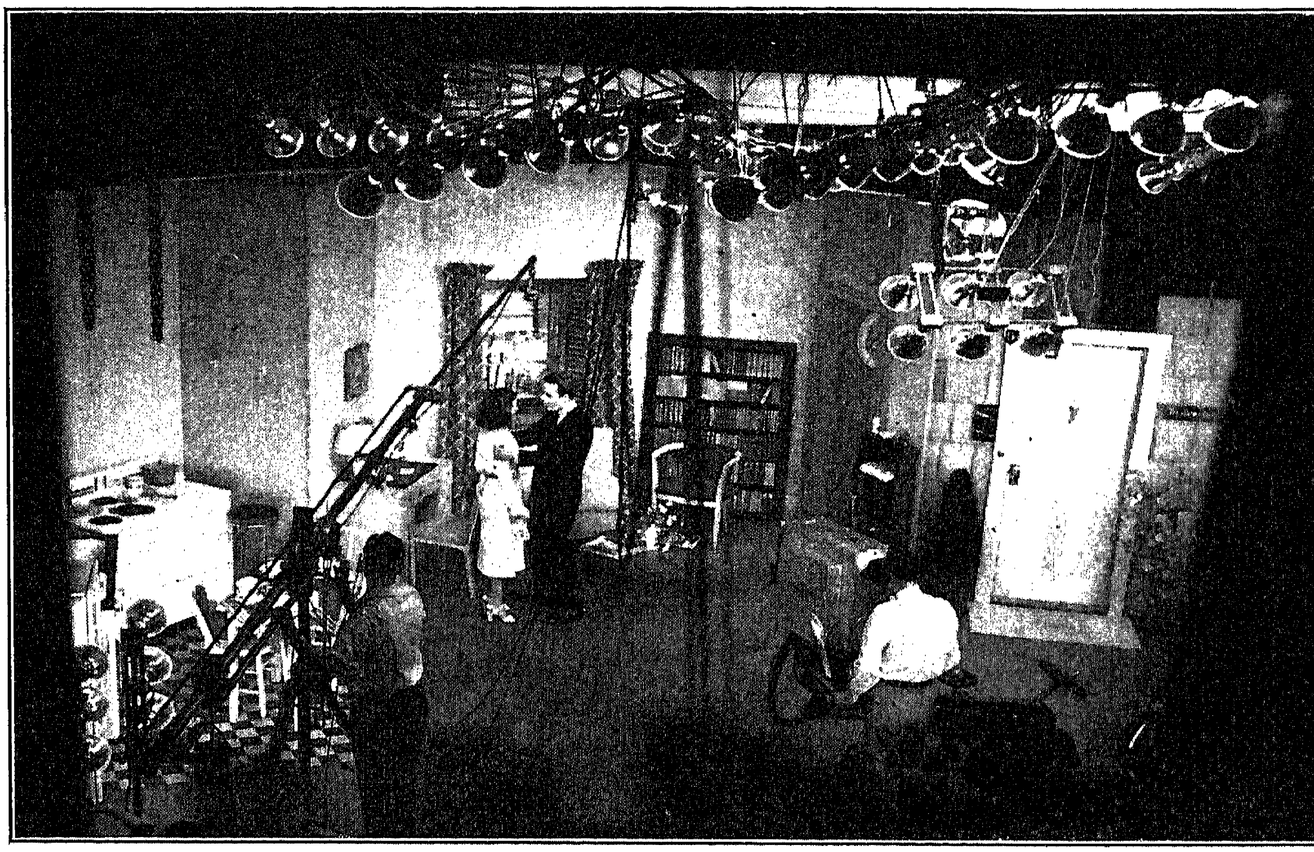


FIG. 18.2.—Lighting Arrangement for Typical Studio Set (courtesy of NBC).

background scenery on a transmission screen located at that end of the studio. This projected background adds greatly to the flexibility of the staging because it permits the use of moving as well as stationary scenery. To eliminate direct optical reflections which might interfere with the projected image, the end wall is painted black. The other three walls are coated with aluminum paint to increase the optical efficiency, and the floor is finished with light gray linoleum.

An important feature of the studio arrangement is the communication system between the operating engineers in the control booth and the staff on the studio floor. The control booth is located at one end of the studio and is raised so that its floor is about 10 feet above that of the studio. A window extending across the end wall of the studio makes the

entire studio visible from the control booth. It is of such a height that operators seated in front of any of the control equipment have an unobstructed view. The window is soundproof to prevent conversation carried on in the booth from being picked up in the studio, and furthermore it is

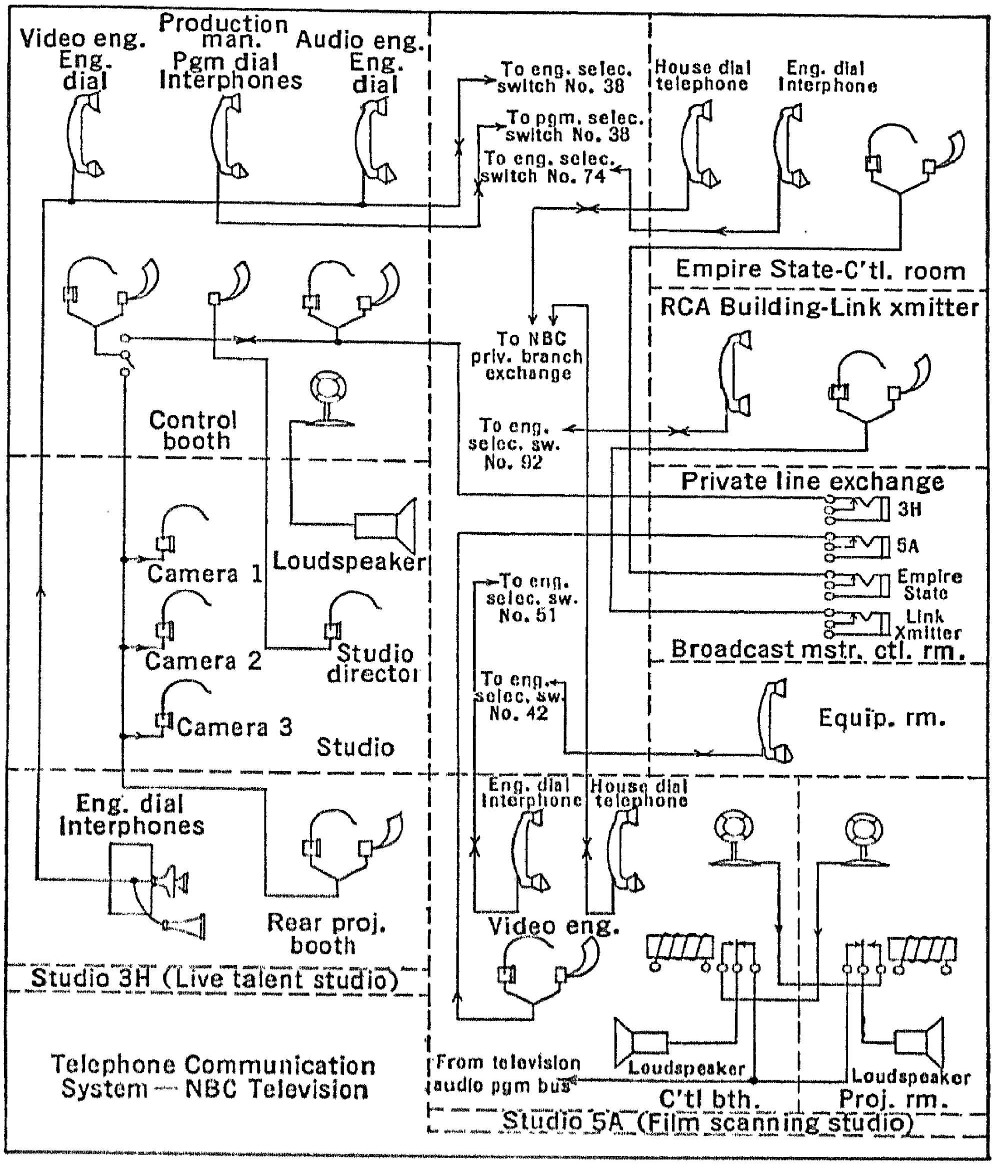


FIG. 18.3.—Schematic Diagram of Communication System Used for Coordinating Television Programs (courtesy of NBC).

partially covered with a filter which attenuates the light from the studio in order to facilitate viewing the monitoring Kinescopes.

A private telephone line interconnects the studio control booth, the film control room, the link transmitter, and the Empire State transmitter, so that during transmission the necessary instructions can be handled. Another line connects the camera operators with the control engineers in the booth, this being a one-way system from the booth to the camera, as any speech on the part of the camera operator might interfere with the

program. A third line allows the production director in the booth to communicate with his assistant in the studio. For general instructions during rehearsals and preparatory to transmission, the studio is equipped with a public-address system operated from the booth. In order to avoid any possibility of accidental use, the public-address system is automatically killed when the studio goes on the air. Communication from the studio to the booth can be carried on over the pickup microphones and monitoring loudspeakers during the time the studio is not on the air. When the studio is on the air, spoken communication cannot be used because of the risk of its interfering with the program. A schematic diagram of the communication system in connection with the television equipment is shown in Fig. 18.3.

18.4. Iconoscope Camera. The cameras carrying the Iconoscopes must be very maneuverable in order to be able to preserve the continuity of the program and yet avoid the stilted appearance which would result from a fixed viewpoint. A camera of the type employed at present is shown in Fig. 18.4. It is mounted on a pedestal made movable by three rubber-tired wheels. The wheels are interconnected by a chain drive and are steered with a lever located halfway up the pedestal. The camera head can be tilted and rotated independently, and it can be locked in any position about either of these axes while left free to move about the other. The "planning" (i.e., panoraming) and tilting are made easy by the careful balance of the camera head. Two handles, resembling the handlebars of a motorcycle, attached to the head, aid in performing these operations. Like the grips on motorcycle handlebars, the ends of the arms rotate. Turning the grip on the right focuses the camera lens; rotating that on the left raises or lowers the camera head. A motor driving a windlass and pulley arrangement operates the telescoping elevating tubes.

The camera operator must be able to follow the action on the stage accurately and keep the camera in constant focus. Therefore an accurate and convenient "finder" is required. For this reason the studio cameras are equipped with a pair of identical lenses, one of which focuses the scene in front of the camera onto the mosaic, while the other focuses it onto a ground glass which is the same distance from this lens as the mosaic from the other. The operator by watching the ground glass of the finder can obtain a sharp image on both the mosaic and the ground glass. The two lenses are mounted on the movable front of the camera, which can be racked in or out by means of the handle mentioned above. This type of finder permits watching the image of the scene being televised with a wide open lens, irrespective of the stop required by the Iconoscope lens, this viewed image having not only the maximum brightness, but also the most critical focus.

Interchangeable pairs of lenses having focal lengths of 6½, 14, and 18 inches are provided for each camera. These lenses are mounted on face plates which can easily and quickly be slipped into place on the front of

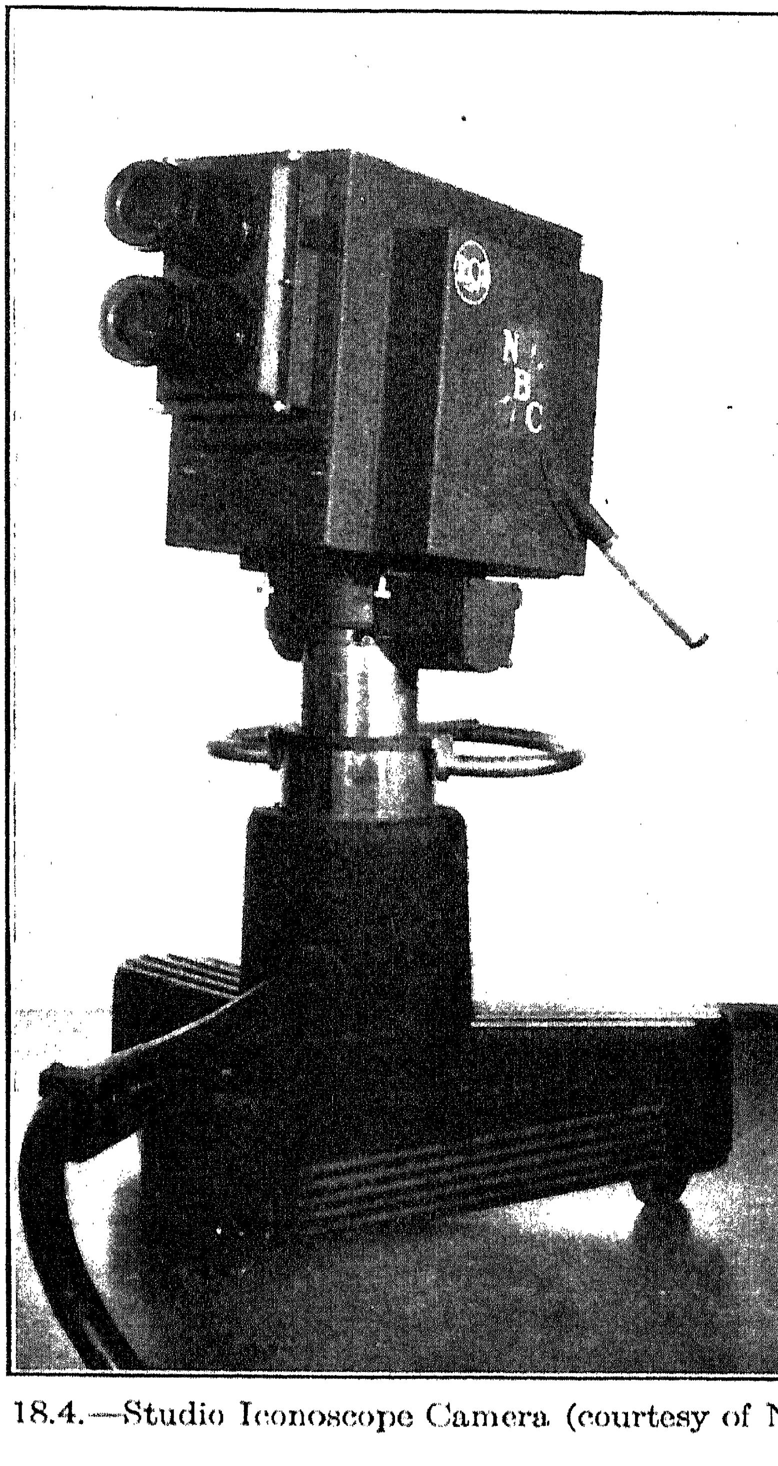


Fig. 18.4.—Studio Iconoscope Camera (courtesy of NBC).

the camera. The short-focal-length lens permits covering an entire set at one time; with the long-focal-length lens a close-up of individual actors can be obtained without moving the camera into the field of another camera using a wide-angle lens.

The cameras can be mounted on moving-picture camera dollies as

illustrated in Fig. 18.5 as an alternative to their usual pedestal. When so mounted they can be moved about the set by an attendant, thus freeing the operating engineer to concentrate his attention on focusing the camera and following the action.

The camera head, which is shown open in Fig. 18.6, not only carries the Iconoscope and its associated deflecting yoke, but also contains a blanking amplifier and a video pre-amplifier. The pre-amplifier consists of a

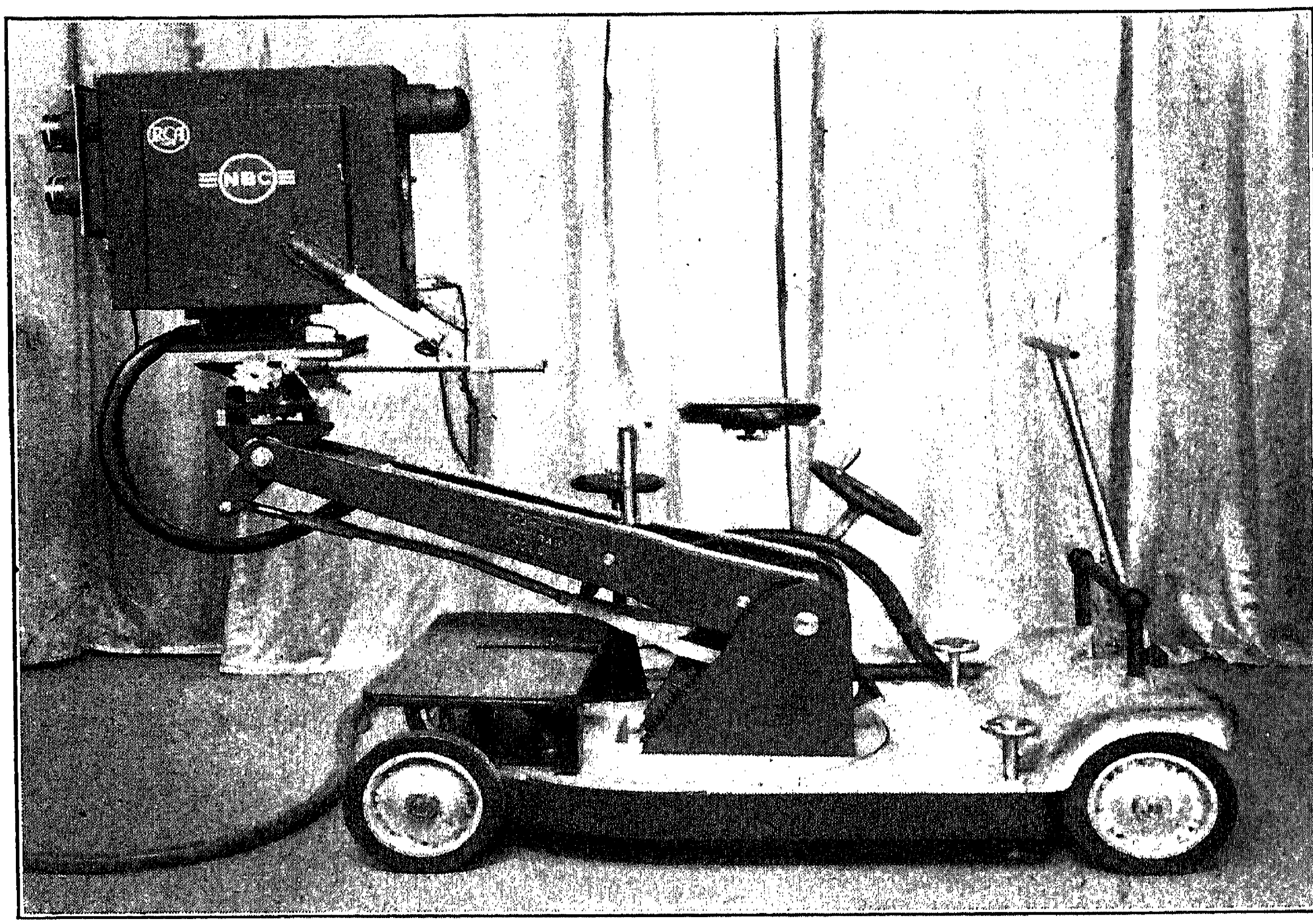


FIG. 18.5.—Iconoscope Camera Mounted on Dolly to Increase Mobility (courtesy of NBC).

cathode follower first stage, two compensated resistance-coupled stages, and a final low-impedance output stage, which is matched to the video cable. A thirty-two-wire cable, approximately 1½ inches in diameter and 60 feet long, connects the camera with the control booth. This cable supplies the constant potentials needed for the Iconoscope, vertical and horizontal deflection, blanking signals, etc., and also includes a low-capacity video cable to carry the picture signal output from the pre-amplifier.

Three of these cameras are used in the Radio City Studio, each being entirely independent of the others as far as the studio control room. In addition to these cameras for use with the conventional Iconoscopes, two

cameras are available for use with special Iconoscopes, thus permitting tests on tubes which are still in the developmental stage.

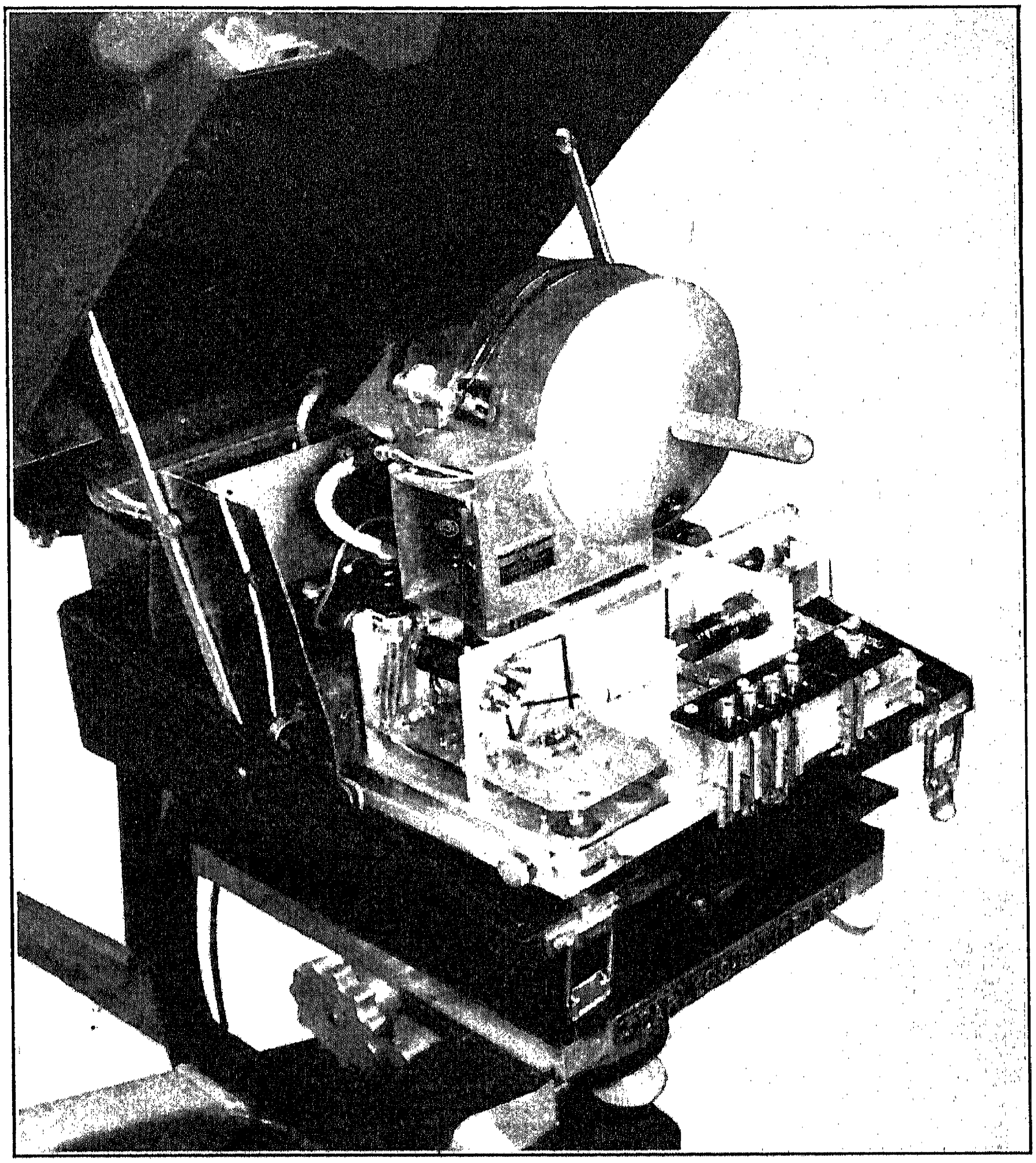


FIG. 18.6.—Interior of Iconoscope Camera (courtesy of NBC).

18.5. Direct Pickup Studio Control Room. Both sound and picture are monitored from the control booth. The sound monitoring is similar to the familiar sound broadcast monitoring, and a discussion of it does not belong in the present pages. The video monitoring equipment and technique are, however, very different and merit rather detailed consideration.

Unlike sound broadcasting practice, the video output from the preamplifiers of the cameras cannot conveniently be mixed and fed into a single amplifier in the control room. Instead a separate amplifier must

be provided for each camera. This is because a number of control operations, differing for each camera, must be performed in the amplifier. Furthermore, it is difficult to switch and feed video circuits at low levels. Consequently three amplifiers corresponding to the three camera circuits must be part of the installation in the control booth. In addition, there are the power supplies and the deflection equipment for each of the three cameras.

Although there are three amplifiers corresponding to the three cameras, these units are completely interchangeable, the cable from each camera terminating in a multi-pin plug which fits into any of three receptacles in the wall of the studio connected with the amplifiers.

The amplifiers are compensated resistance coupled, of the type described heretofore, and work at an input level of about 0.1 volt, while their output is of the order of 1.0 volt. Both input and output impedances are low, the former to match the camera cable, the latter to permit feeding a concentric cable to the line amplifier.

Each amplifier is equipped with a gain control, and a brightness or background control which regulates the pedestal height between lines, giving the d-c level.

Also associated with each amplifier is a shading circuit. These circuits provide the wave form required to compensate for the spurious signal generated by the Iconoscope. Since the wave form of the spurious signal differs for different Iconoscopes, and varies with the particular operating conditions, the compensating wave form must be under the control of the operating engineer. The compensation is accomplished by generating the various components appearing in the final shading signal, adjusting the amplitudes of these components independently by means of suitable control knobs, and mixing the results to produce the output wave form. The simplest circuit uses two horizontal and two vertical wave components, one of these components being a sawtooth wave, the other parabolic. These can be added with any amplitude or sign. Another shading circuit makes use of sinusoidally shaped components providing half a sine wave, a full sine wave and three-halves of a sine wave. These components can be controlled in amplitude and phase. A shading circuit of the first-mentioned form is used at Radio City. The output from the shading circuit after being amplified is added to the picture signal by means of conventional injector circuits in the video amplifier.

The controls for the various functions mentioned or implied must be so placed that they can be conveniently used by the operating engineer. These controls can be classified into two groups, namely, those requiring constant attention during the program production, and those which are used only for initial adjustments before the start of the program. The

placing of the former obviously takes precedence over the latter, as they must be within easy reach of the operating engineer seated in the monitoring position. This first group of controls includes gain and brightness controls, shading controls, and the controls for switching the cameras. The second group consists of those controls associated with the scanning, that is, amplitude, centering, and keystone adjustments for vertical and horizontal Iconoscope deflection, and also the controls regulating the

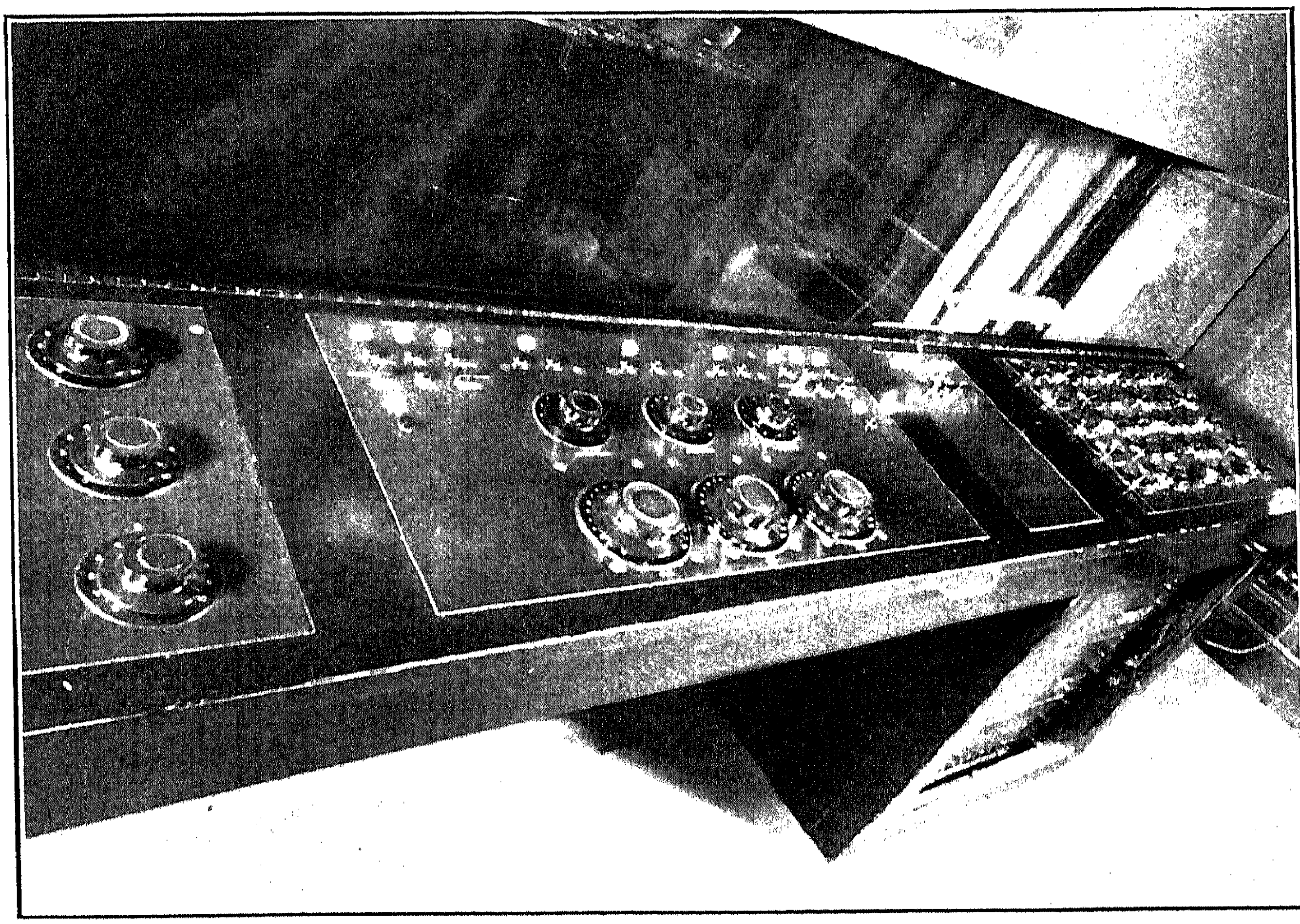


FIG. 18.7.—Studio Control Console (courtesy of NBC).

focus and beam current of the gun, together with a number of others of a similar nature.

The actual arrangement of the first group of controls is shown in Fig. 18.7, illustrating the control console, or desk. The three large knobs arranged in a horizontal row seen at the center of the picture are the gain controls for the three amplifiers. Above them are three slightly smaller knobs which regulate the pedestal. The row of push-button switches controls the relays switching the cameras, which will be described later. At the extreme right can be seen three small panels each containing ten knobs and six switches. These are to regulate the amplitude of

the shading components for each amplifier. The three vertically arranged knobs seen at the left control the beam focus for the three Iconoscopes.

Not unrelated to the location of the controls is the location of the equipment itself. In the case of the shading circuits and shading amplifiers, where long output leads do relatively little harm, the equipment can be located directly under the panel carrying the knobs. The gain controls cannot be placed near the video amplifiers, however, without a

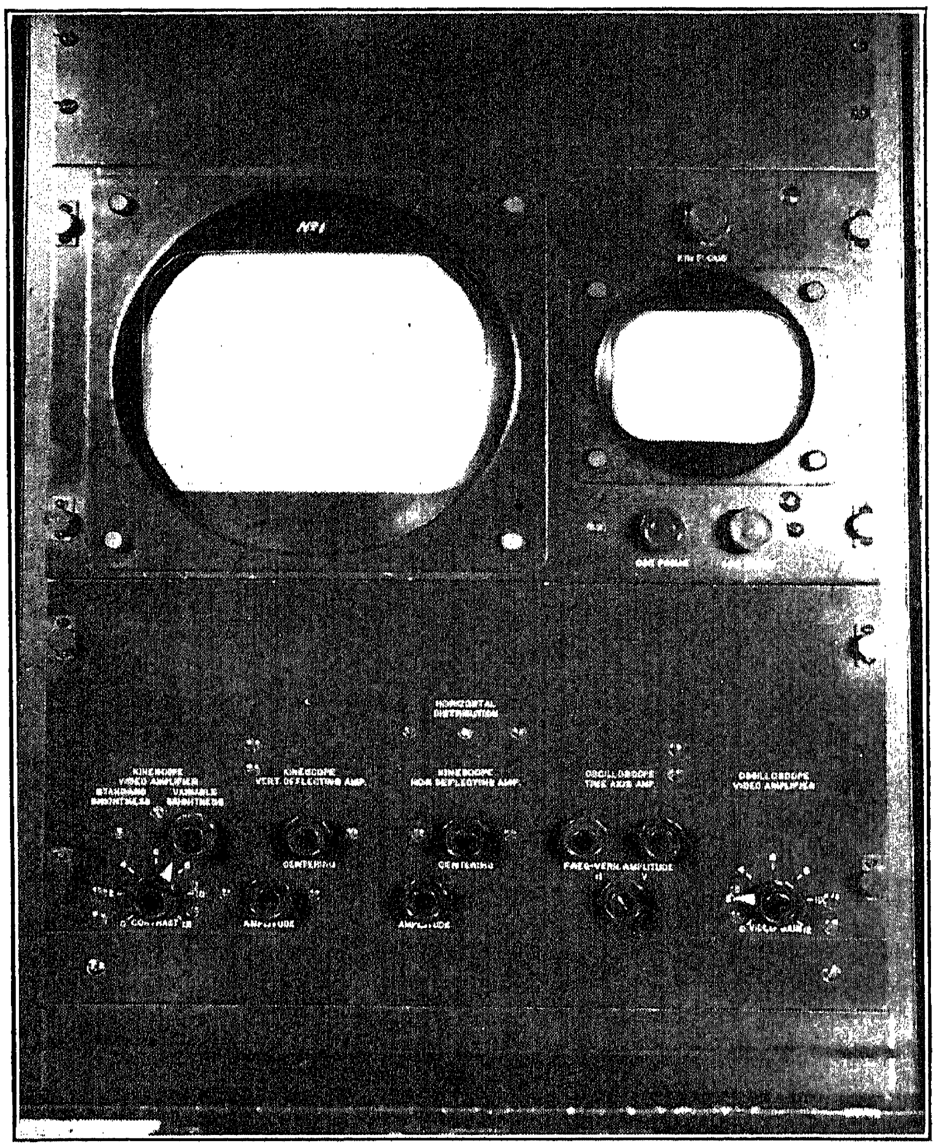


FIG. 18.8.—Monitoring Unit (courtesy of NBC).

considerable sacrifice in operating convenience. On the other hand, long leads to the gain control might seriously impair the response characteristics of the amplifier. One solution for this problem is the use of Selsyn motors. The control knob is geared to the armature of one of a pair of Selsyn motors, while the armature of the other is geared to the gain potentiometer located at the optimum physical position in the amplifier.

Any rotation of the control knob causes a corresponding rotation in the armature of the motor attached to the potentiometer. This type of control has been tested and found very satisfactory. Another way of accomplishing the same end is by one or more stages of amplification which are so designed that the gain can be controlled through the grid bias. This second method is the one actually employed at this time.

In the control booth, at locations convenient for continuous observation by the control engineer, are two 9-inch monitoring Kinescopes. Accompanying each Kinescope is a 5-inch oscilloscope tube which shows the wave form of the video voltage. Fig. 18.8 shows one of these monitor-groups. Each Kinescope has its own picture amplifier located back of the panel carrying the tube, which permits operating on a minimum video level of approximately 0.5 volt peak to peak. The switching is so arranged that either of these monitors can be connected to the output of any one of the three camera amplifiers, or can be connected to the output of the line amplifiers feeding the links which connect Radio City with the Empire State Building transmitter. This switching arrangement will be considered more in detail later. In normal operation, one monitor will be showing the output of the line amplifier, that is the actual picture that is being sent to the transmitter, while the other monitor will be connected to the amplifier of the camera which is next to be used, in order to insure that it is properly placed and adjusted.

The deflection generators with their associated controls, the power supplies for the Iconoscope cameras, and the Iconoscope blanking amplifier are located in racks behind the operating desk. Ample space is allowed between the rear wall of the booth and the racks to permit easy servicing.

18.6. The Film Studio. The program from the live-talent studio is augmented by the program from the film studio. These two should not function as separate entities but rather as a single unit. To make this possible, accurate communication and switching between the two studios must be maintained, as well as complete control of the equipment in each studio. Only in this way can a single coordinated program be maintained from these two points.

The film studio consists of a projection room and a control room containing the cameras and monitors. In addition there is a film vault, completely fireproofed, and equipped with protective sprinklers not only in the vault itself but also in the individually fireproofed storage cabinets. A diagram of the layout of the studio is shown in Fig. 18.9. The projection room contains two special 35-mm projectors, two 16-mm machines, and a lantern for projecting still pictures and test slides. The two 35-mm projectors are equipped with soundheads, and are arranged to operate at the

standard rate of 24 frames per second, thus permitting the use of any standard moving-picture film. A special synchronized shutter which flashes the image from the film onto the mosaic only during the return time of the beam makes possible the pickup from a film having a speed of 24 frames per second, by an Iconoscope using interlaced scanning with 30 pictures per second. The mechanism of this form of pickup was described in Chapter 10. The small 16-mm projectors are not arranged for

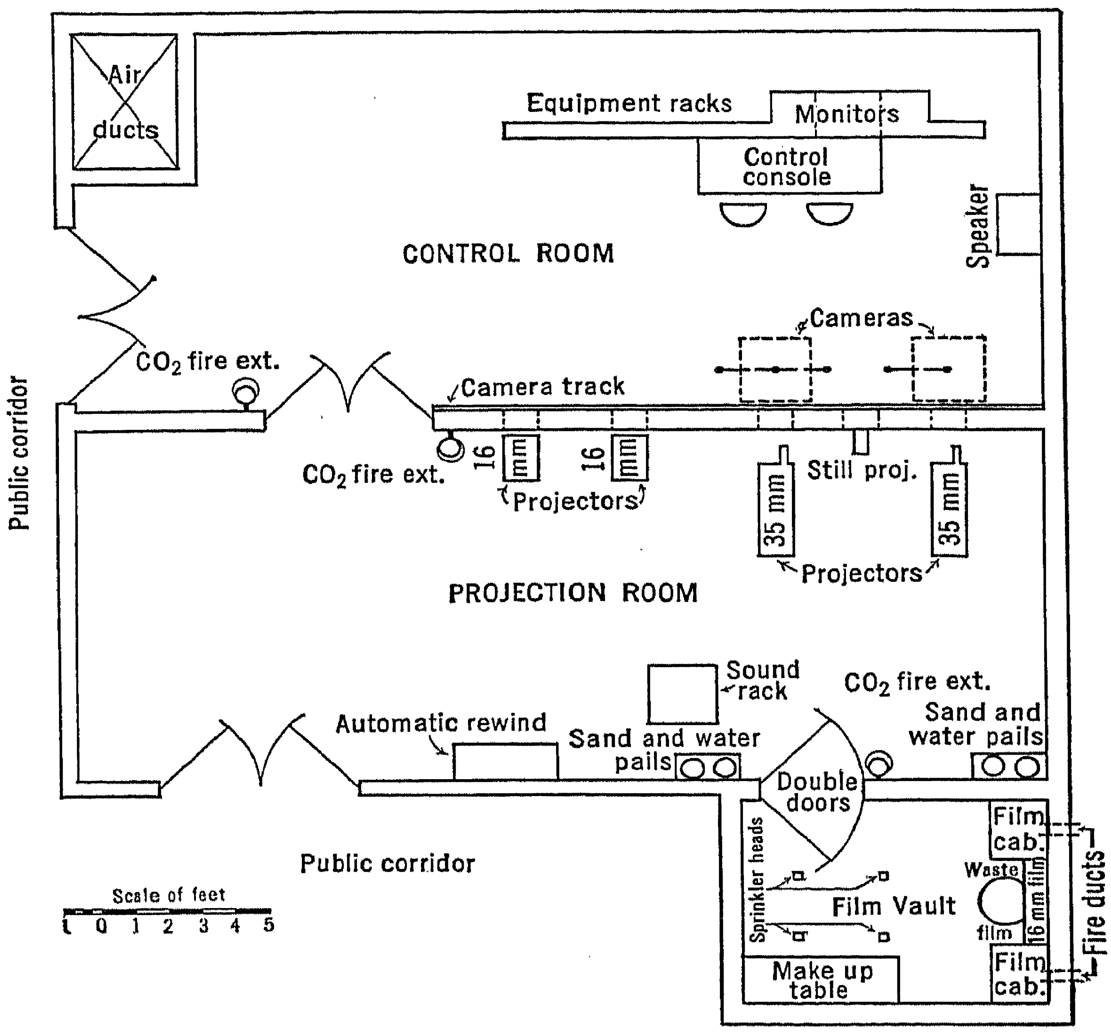


FIG. 18.9.—Plan of Film Studio (courtesy of NBC).

sound, and run at 30 frames per second. Oversized synchronous motors are necessary to power this special projection equipment. In order to synchronize the film speed with the video system the motors are driven with alternating current from the supply which feeds the synchronizing equipment of the television system. Fig. 18.10 is a photograph of the projection room.

Between the projectors and the Iconoscope cameras is a fireproof wall, having five protected windows through which the pictures are projected. The cameras are located in the control room containing the monitoring

equipment. Two cameras are provided, and are arranged to move on a track so that either camera can be placed before any of the four projectors. Like the live-talent studio cameras, these cameras contain video pre-amplifiers and also blanking amplifiers.

The control equipment associated with these cameras is very much the same as that in the previously described control booth, both with regard to apparatus and physical layout, although of course only two sets of controls are necessary. Two monitoring Kinescopes, each with an associ-

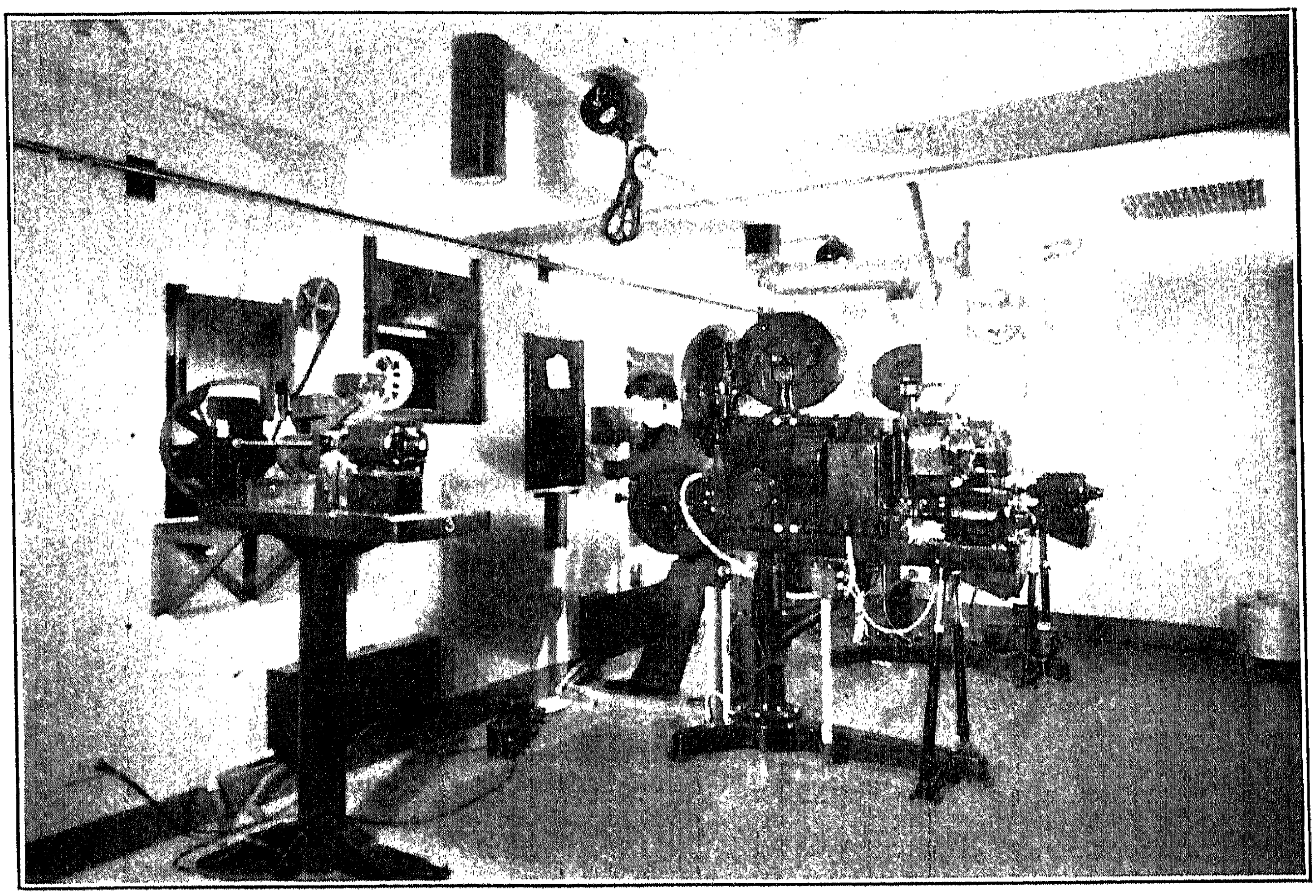


FIG. 18.10.—Moving-Picture Units Used for Film Transmissions (courtesy of NBC).

ated oscilloscope, are provided. The film control console is similar to the one just described but contains only two pedestal and gain knobs and two banks of shading controls. An overall view of the control equipment for this studio, including both sound and picture monitors, is given in Fig. 18.11. The racks in the left foreground house the deflection generators and Iconoscope power supplies; next to them are the panels containing the sound amplifiers and controls. Beyond these, in the center background, is the video channel. The control console showing an operating engineer seated at the monitoring position occupies the center of the picture. At the extreme right can be seen an Iconoscope camera.

Like the direct pickup studio, the film studio is equipped with a com-

plete communication system. Between the projection room and the control station there is a two-way microphone-loudspeaker system. This is used instead of the conventional telephone because it is often necessary, when switching projectors or moving cameras, for more than one engineer to receive the complete instructions. Except when operated as an address system the two loudspeakers are used for monitoring. Also there is a wire from the control console which is connected with the live-talent studio,

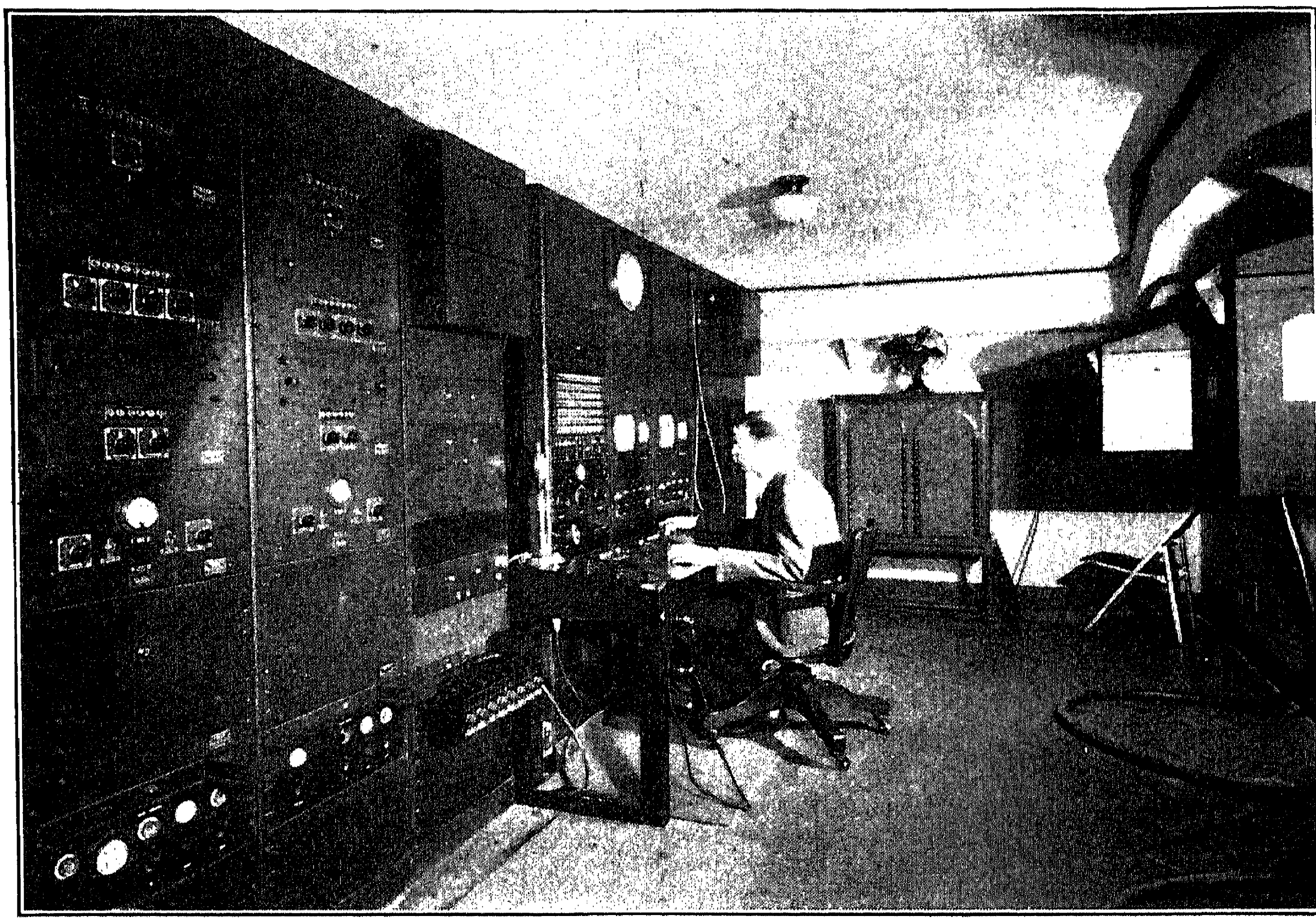


FIG. 18.11.—Camera and Control Room for Film Pickup (courtesy of NBC).

the link transmitter, and the Empire State transmitter. In addition there are the building and the engineering dial telephones.

18.7. The Control Equipment Room. The line amplifiers and synchronization generators which are common to both studios are installed in a main equipment room. Two views showing the front and back of the main racks in the equipment room are given in Figs. 18.12 and 18.13.

The output of the video amplifiers in the two studio control rooms may be fed into any of three line amplifiers. These amplifiers are operated at an input level of about 1 volt and have an output of the order of 10 volts with an output impedance to match a 75-ohm coaxial line. This output may be connected to the cable which leads directly to the Empire State transmitter, or switched to a line leading to the radio relay link.

The synchronization generators produce the blanking and deflection impulses for both studios and also the synchronizing impulse which is fed into the line amplifiers to become that part of the transmitted video signal which locks the receivers into synchronism with the transmitter, as has already been explained. There are two identical synchronization generators, either of which may be used, thus insuring that no breaks will occur in the program due to failure of this essential equipment. In

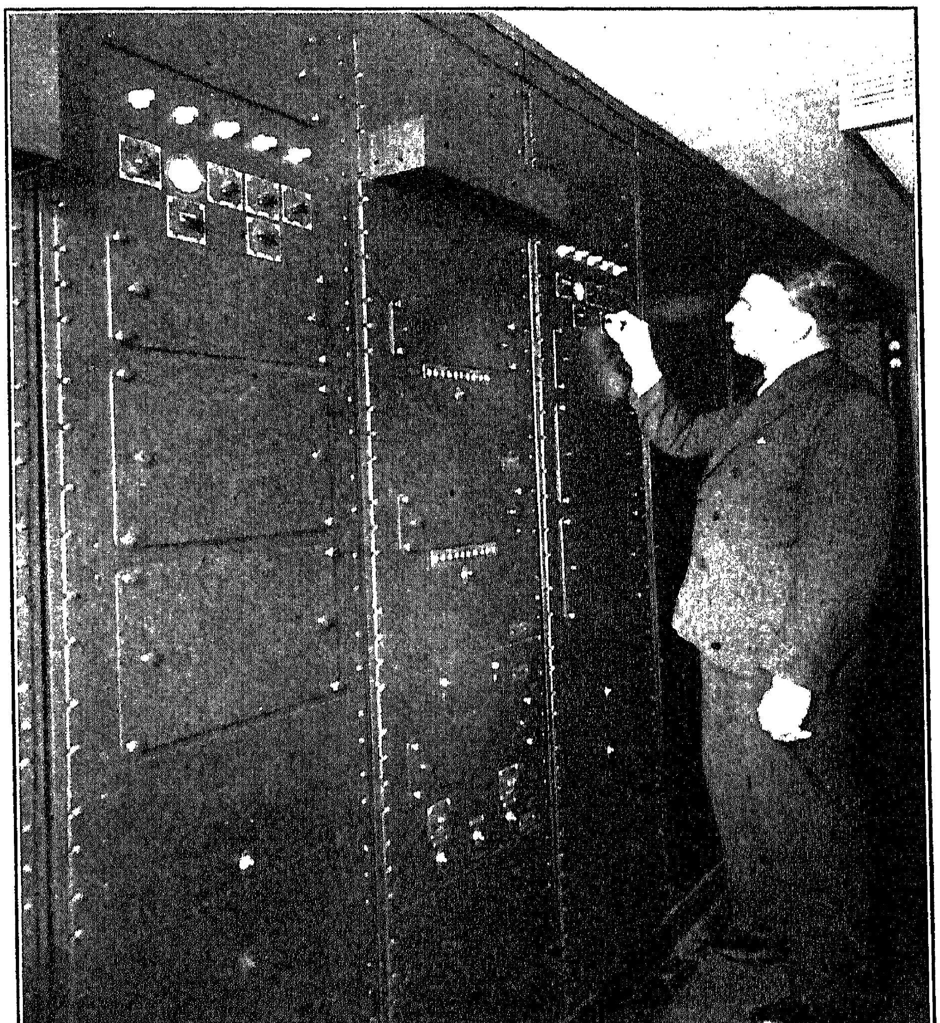


Fig. 18.12.—Synchronizing Generator and Video Line Amplifier Panels (courtesy of NBC).

Fig. 18.13 one of these generators can be seen at the extreme left. Next to it, and separating it from the other generator, is the amplifier which amplifies the output of the generators prior to its distribution. At the right of the photograph the three line amplifiers are visible.

The synchronization generators are locked into the 60-cycle outside power line supplying the building, so that the television system is essentially synchronized with the power lines which supply many of the receivers within the service area of the transmitter. In this way any pos-

sible interaction with the alternating current at the receivers (i.e., hum) is rendered as inconspicuous as possible. It also makes it possible to operate the film projectors with synchronous motors directly on the normal

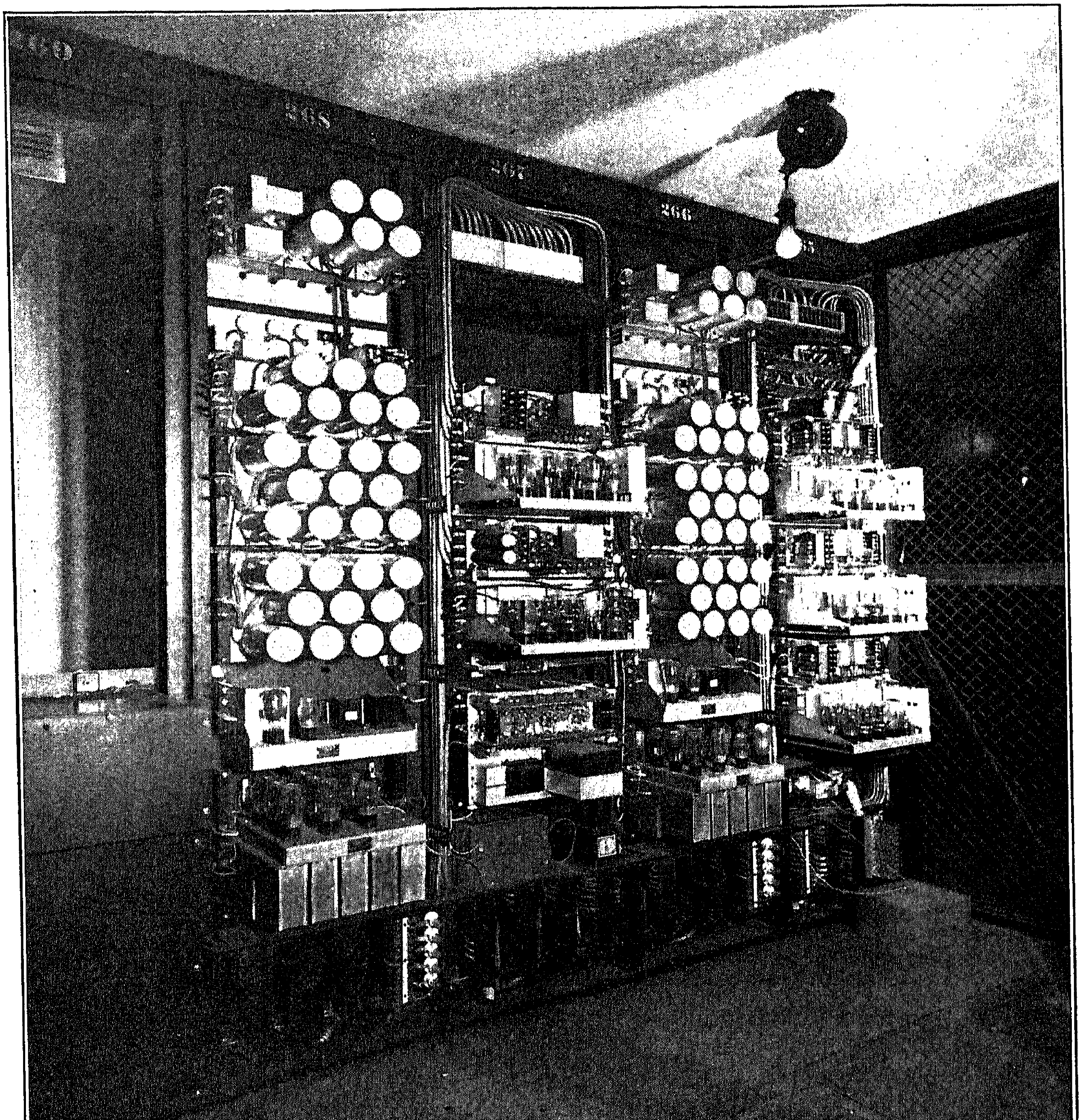


FIG. 18.13.—Rear View of Synchronizing and Video Line Amplifier Panels (courtesy of NBC).

power supply. To insure maximum stability, the generators are operated continuously twenty-four hours a day. By means of suitable relays, either generator may be instantaneously switched to the coaxial feeders supplying the studios, the other generator acting as a standby.

18.8. Video Channel Switching. The problem of switching video channels is not easily solved. It must be extremely rapid if interruption of the program is to be avoided. Also it must be done in such a way that it does not sacrifice any picture quality, either by causing loss of resolution as a result of attenuating the high frequency, or by introducing phase or amplitude irregularities such as would produce transients in the picture, or by permitting crosstalk. Finally, it must be carried out without introducing surges which might endanger vertical synchronization.

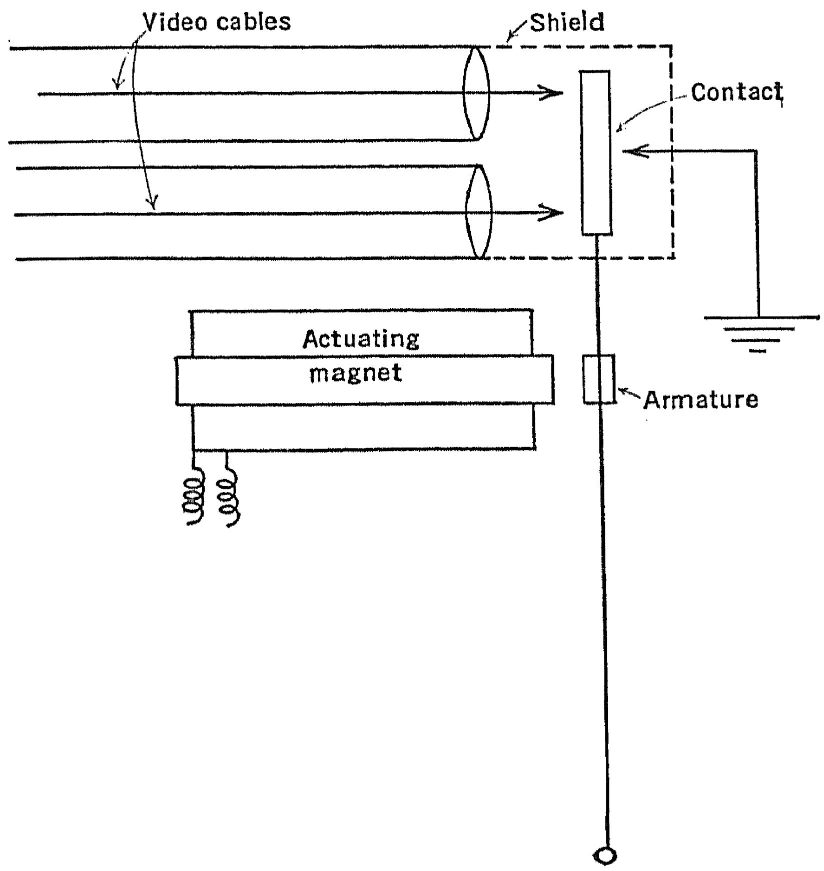


FIG. 18.14.—Schematic Diagram of Remote-Control Video Switch.

The switching is done entirely by remotely controlled relays located in the optimum position with respect to the elements being switched. The relays are of the spring contact type, the contact being made as small as is consistent with mechanical reliability and positive action, in order to minimize capacity to ground. Two moving contacts in series are used in all relays so that when the relay is in the open position these contacts are grounded, serving as a shield between the open end of the incoming and outgoing video cables. An explanatory diagram is shown in Fig. 18.14.

The relays switching the camera in each studio are interlocked so that by pressing the push button assigned to any given camera it is put into the circuit and any other camera which may have been in this circuit is dropped. This insures smooth, rapid camera switching. In order to do

this without introducing a surge into the circuit, the sequence and the timing of the switching relays must be very accurately adjusted.

Fig. 18.15 shows schematically the relay switching at the Radio City installation. It will be seen that in either studio any monitor can be operated from any camera irrespective of whether the latter is in the line amplifier circuit. Any camera, but only one camera, may be connected to any one of the three line amplifiers. If a film camera is connected to the line amplifier, a camera push button in the direct pickup studio control room can cause it to be dropped and that camera to be inserted. The controls are arranged, however, in such a way that the film studio controls cannot drop the direct studio camera.

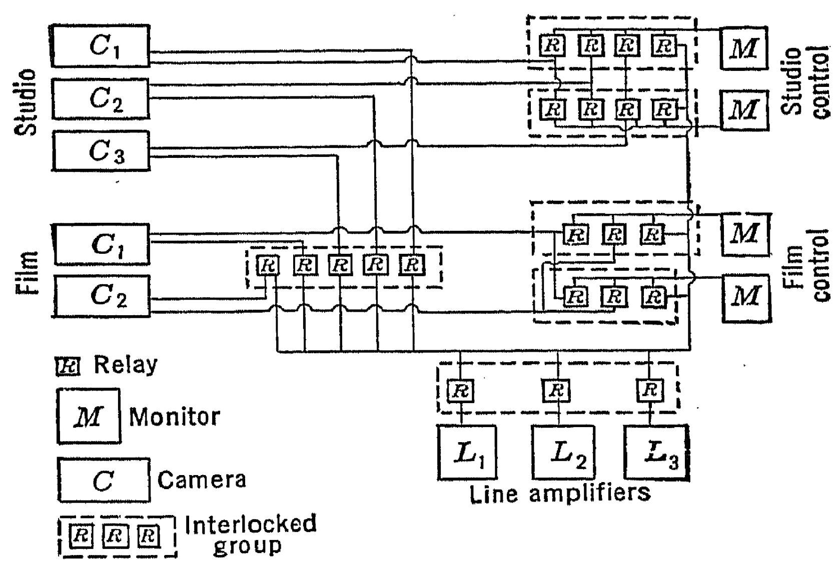


FIG. 18.15.—Television Relay Switching System.

Monitors in both the studios can be switched to the output of the line amplifier and are generally so connected when a program is being transmitted.

18.9. Special Problems. There were many important problems which had to be solved in engineering an installation such as that of the Radio City television system. It is manifestly impossible to discuss even an appreciable fraction of them in the space here available. Two or three might profitably be mentioned as illustrative of the type of problem encountered in a project of this nature.

The plate supply for the resistance-coupled amplifier has its source in a large central storage battery. The amplifiers were connected to this battery by low resistance lines, in some instances having a nominal length of 100 or 200 feet. The reactive impedance of these lines was found to be sufficient to introduce serious non-uniformities in the amplifier response, if not to cause actual instability and oscillation. Various bypass con-

densers were tested in order to decrease this line impedance. It was found that, in order to reduce the impedance sufficiently over the working band of the amplifiers, a capacity of 1000 microfarads was necessary.

All video signals are transmitted over single low-loss coaxial cables, having a surge impedance of about 76 ohms. This means that the amplifier loads are unsymmetrical with respect to ground; therefore single-ended rather than push-pull amplifiers are employed throughout. Initial tests proved the feasibility of the arrangement and indicated that inter-

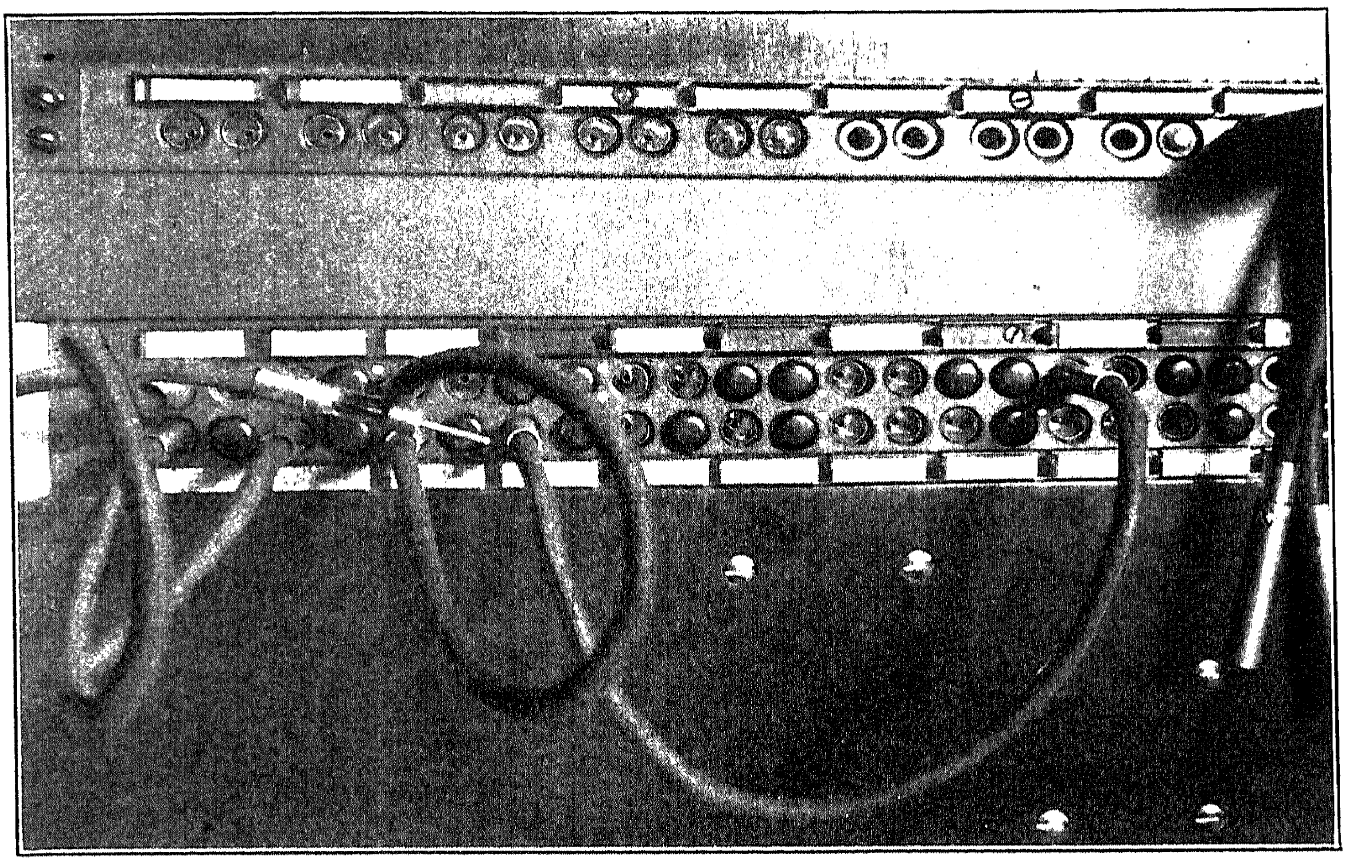


FIG. 18.16.—Coaxial Cable Jack Panel and Patch Cords (courtesy of NBC).

ference due to currents in the outer cable sheath was negligible, thus making double cables unnecessary.

All cable terminal blocks are completely shielded with heavy copper partitions, to insure against impedance irregularities and crosstalk. Frequently it is desirable to provide means for making changes in the connections between various elements of the video system, where these changes are not necessary frequently enough to warrant relays. For this purpose special coaxial jacks and patch cords were designed, which permit rapid changes with no sacrifice in response characteristics. A coaxial jack panel is illustrated in Fig. 18.16.

Safety precautions were taken throughout, errors, where they occurred, being on the side of conservatism. Besides the normal fire precautions,

including sprinkler systems, fireproofed walls, and a complete alarm system, every possible guard against electrical hazards has been taken. All covers and doors exposing leads having voltages over 500 volts are interlocked, so that when they are opened the power is automatically removed. All doors on the studio cameras are manually locked as well as interlocked, so that the proper keys are required to open them. When work must be done on equipment in the racks, grounding hooks and rubber mats are used. Furthermore, engineers are not permitted to work singly on any apparatus which could, through failure of equipment, negligence, or any other cause, become dangerous.

A great deal of very valuable information on television technique has been gained through the tests which have been made on this installation. The installation has ceased to be experimental, and has become a practical broadcasting unit, not only as far as equipment is concerned, but also from the operating standpoint. The Radio City plant has now reached a point where programs of high entertainment value and interest, smoothly transmitted, are reliably assured.

18.10. Test Receivers. The receivers used in the television field tests provided data not only on the quality of the transmission and programs but also on the feasibility of a television receiver as a home instrument. During the course of the investigation many changes were necessary in these receivers in order to keep abreast of the advances in the field, and many ways were found to improve the receivers; consequently the models used during the final year of the test were capable of a much better picture than the earlier sets. One of the later models is illustrated in Fig. 18.17. When the lid is closed the set resembles a console broadcast receiver. On the inside of the lid is a large mirror, so arranged that when the lid is open the picture can be seen reflected in it. This viewing arrangement makes it possible to watch the picture from any convenient position in front of the set and at the same time shields the face of the Kinescope from the room lights.

Near the base of the cabinet can be seen the grille covering the loudspeaker which provides the accompanying sound.

Seven knobs are provided on the front of the receiver to control the audio and video performance. The main tuning control for picture and sound is in the center and covers a frequency range from 40 to 100 megacycles. The three knobs seen at the right are for the volume and tone of the sound. Contrast, detail, and background brightness are governed by the three controls on the left. Although the seven controls make the operation of the set somewhat more complicated than that of an ordinary broadcast receiver, yet the controls are logical, performing separate,

easily recognizable functions, so that an operator who is totally unfamiliar with the theory of television should experience no difficulty.

The internal arrangement of the receiver can be seen from Fig. 18.18. The Kinescope is mounted vertically with its viewing screen directly

Fig. 18.17.—An Experimental Console-Type Television Receiver.

under the lid. Below it is the voltage supply for the Kinescope and for the audio and video circuits. The receiver chassis is mounted against the front wall of the cabinet, and includes the radio-frequency circuits, the intermediate audio and video amplifiers, and the deflection generators.

A circuit similar to those described in Chapter 17 is used. The ultra-high-frequency input is not amplified ahead of the first detector. There are two intermediate amplifiers, one for audio and one for video signals, the signal separation being made in the output of the first detector. The selectivity of the intermediate amplifier is so adjusted that only the carrier and one sideband are amplified. At the video second detector the signal is separated into its video and synchronizing components. The

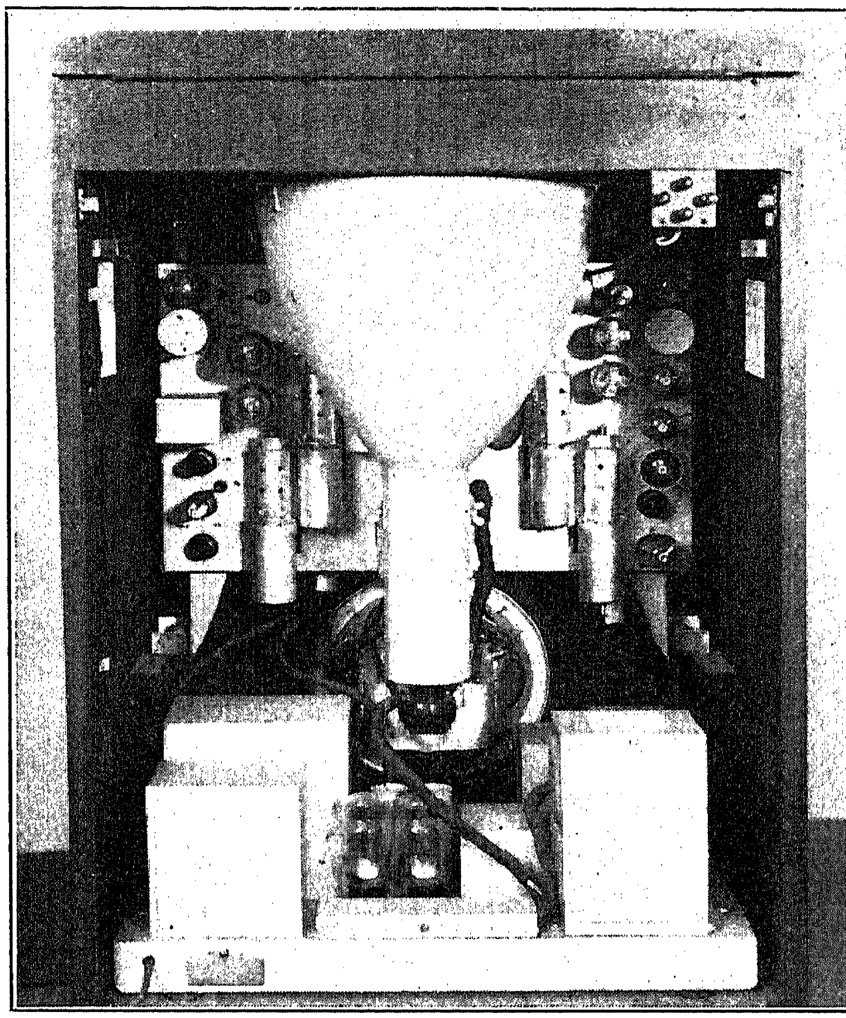


FIG. 18.18.—Rear View of Experimental Receiver.

latter controls the vertical and horizontal sawtooth deflection generators, the deflection being magnetic in both directions. The video signal after being further amplified and having the d-c component reinserted is supplied to the control grid of the viewing tube. Automatic gain controls are provided for both the audio and video circuits.

As has been mentioned, three controls permit adjusting the picture quality. The contrast control regulates the magnitude of the video signal applied to the Kinescope—in other words, the overall gain of the amplifier. Detail control is effected by adjusting the response of the amplifier

to the upper frequencies. Finally, the background brightness is controlled by varying the mean bias of the Kinescope.

The practical performance of this type of receiver was found to be excellent. During the course of the tests it was found not only to give a picture of the desired quality, but also to have a long life free from trouble.

These, together with other models used in the tests, were definitely proved suitable to form the basis of further commercial developments.

REFERENCES

1. L. M. Clement and E. W. Engstrom, "R. C. A. Television Field Tests," *R. C. A. Rev.*, Vol. 1, pp. 32–40, July, 1936.
2. R. R. Beal, "Equipment Used in the Current R. C. A. Television Field Tests," *R. C. A. Rev.*, Vol. 1, pp. 36–48, January, 1937.
3. O. B. Hanson, "Experimental Studio Facilities for Television," *R. C. A. Rev.*, Vol. 1, pp. 3–17, April, 1937.
4. R. M. Morris and R. E. Shelby, "Television Studio Design," *R. C. A. Rev.*, Vol. 2, pp. 14–29, July, 1937.

CHAPTER 19

EMPIRE STATE TRANSMITTER

The programs originating from Radio City are transmitted from the Empire State Building. The transmitter is located on the eighty-fifth floor and the antenna on the mooring mast atop the building, 250 feet above the transmitter and 1250 feet above the street.

The transmitter consists of two units, one for sound and the other for the video signal. Originally these two units fed a single coaxial transmission line leading to the antenna which served to radiate both carriers. This antenna consisted of three stacks of horizontal dipoles arranged in an equilateral triangle. Coaxial filters at the points where the transmitters joined the transmission line prevented interaction between the two elements. Double sideband for both audio and video transmission was used, located in the frequency spectrum as shown in Fig. 19.1*a*.

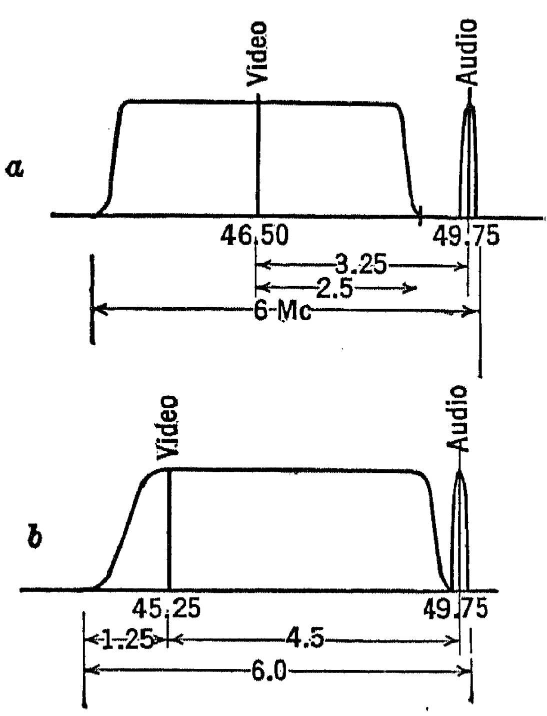


FIG. 19.1.—Single and Double Sideband Frequency Distributions Used in Television Field Tests.

Early in 1939 the system was changed over to single-sideband operation for the video signal, with the frequency distribution given in Fig. 19.1*b*. At the same time separate transmission lines and antennas were installed for sound and picture. A coaxial filter in the transmission line removes the unwanted video sideband. The new antennas and coaxial filters will be described in detail later.

19.1. Location and Arrangement of Transmitting Equipment. In locating the transmitting equipment it was obviously desirable to place it as near the antenna as possible. However, due account had to be taken of the structural limitations of the building. Since the total weight of the two units is nearly 35 tons, it could not be located in the tower directly under the antenna, as this would not have allowed a sufficient factor of safety to comply with the building laws. Even on the eighty-fifth floor

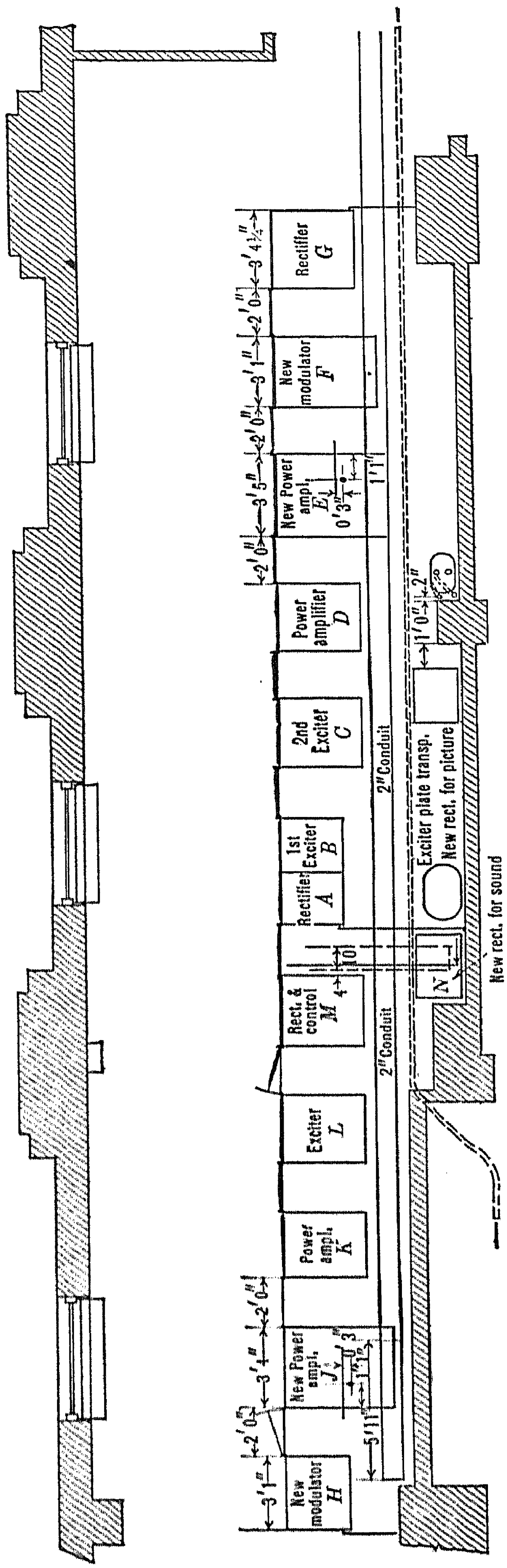


Fig. 19.2.—Layout of Empire State Television Transmitter (courtesy of NBC).

the installation had to be spread over a considerable area to bring the floor loading per square foot within that permitted by good building practice. This, however, could be done without any compromise in efficiency and with but a slight sacrifice in convenience. The transmitter occupies a single large room, whose length is approximately 75 feet and width 20 feet. The general layout of the installation is diagrammed in Fig. 19.2.

The ambient temperature of the transmitter room is maintained at a reasonable value by means of an air-cooling system. Furthermore, in order to reduce dust deposits, the circulated air is washed. Of course ample provision was made for the water necessary to cool the transmitting tubes, some 125 gallons per minute being required for this purpose. The cooling water is arranged so that it can be circulated through cooling radiators, in a closed system, or should it be necessary, so that it can be obtained directly from the city mains.

All possible precautions against danger from fire and electrical hazard have been taken. The room is protected by an alarm system operated by the rate of rise of temperature. A number of hand-operated carbon dioxide fire extinguishers are provided. To eliminate the danger from the high-voltage electrical power, interlocked doors and covers which automatically cut off the power when opened give the only access to points where high-tension leads or electrodes are exposed. In addition, grounding hooks are used when any work is being done on the equipment.

The transformers and motor generators that supply the power required for the transmitter are located in adjoining rooms on the eighty-fifth floor of the building. These occupy a space about equal to the transmitter room. The power for the transformers is brought in through several commercial 13,000-volt feeders, to a stabilizing transformer network on the eighty-fourth floor. Two three-phase four-wire systems supply the transmitters. Each has a separate supply so that they may be operated independently.

19.2. The Audio Unit. The audio unit is a plate-modulated ultra-high-frequency transmitter. A diagram of the equipment is shown in Fig. 19.3. The carrier generator makes use of a crystal-controlled oscillator operating at 3.109 megacycles. The crystal used is so cut as to have little temperature variation and, to further maintain its stability, is contained in a temperature-controlled cell. Four harmonic generators each multiplying the frequency by a factor of 2 bring the carrier frequency to the required value. They are followed by four amplifier stages which raise the carrier power to the level required to operate the final power amplifier stage. This final variable gain stage consists of a pair of plate-modulated RCA 899's operating in a balanced r-f circuit, together with four RCA

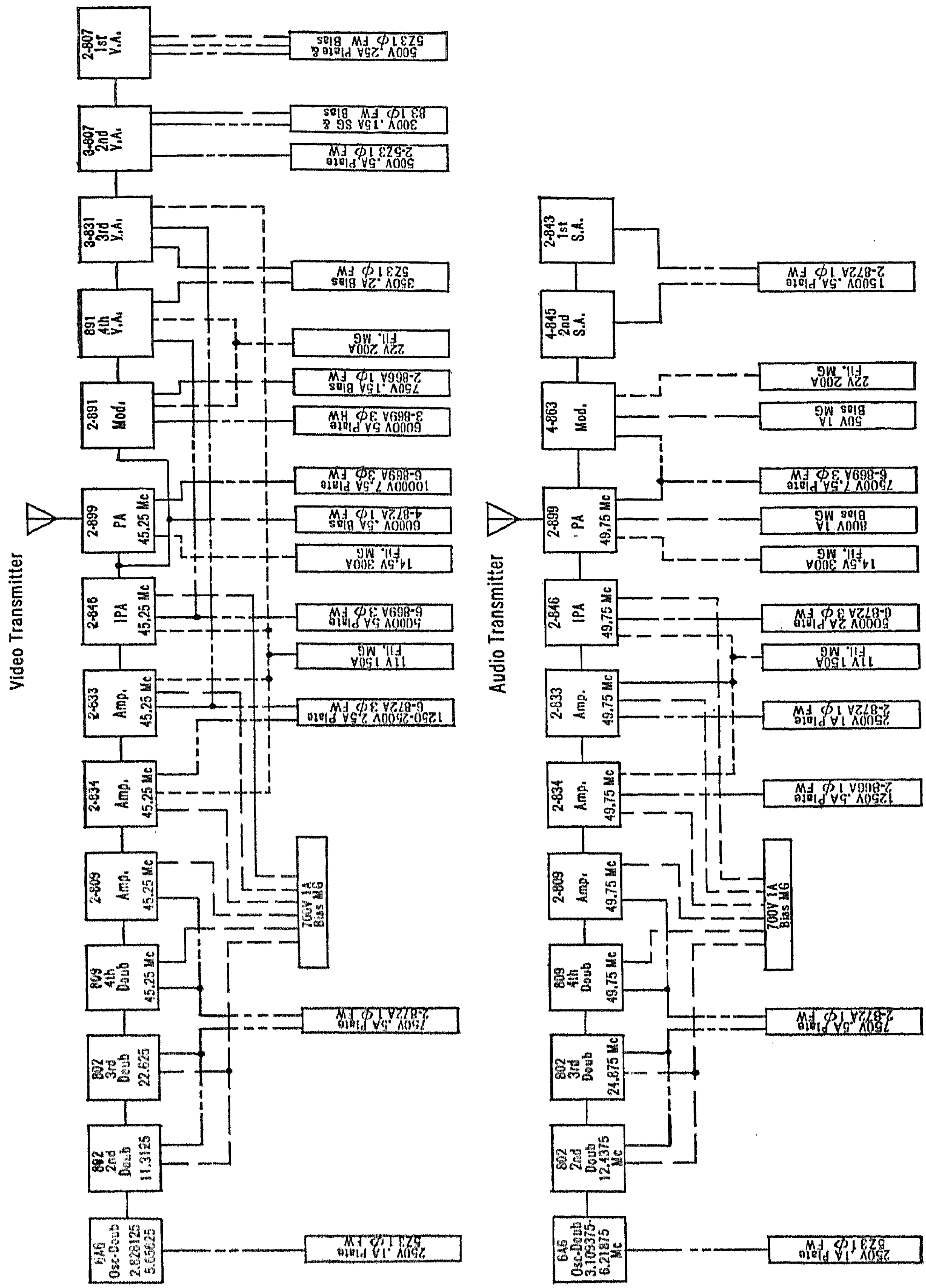


FIG. 19.3.—Block Diagram of Empire State Transmitter (courtesy of NBC).

863's which share a common plate impedance. The four modulator tubes are run class B, while r-f tubes are operated class C with the modulated output coupled directly to the antenna feeder, through a suitable matching transformer.

19.3. The Video Unit. The video transmitter employs an identical carrier generator except that the crystal frequency is 2.828 megacycles.

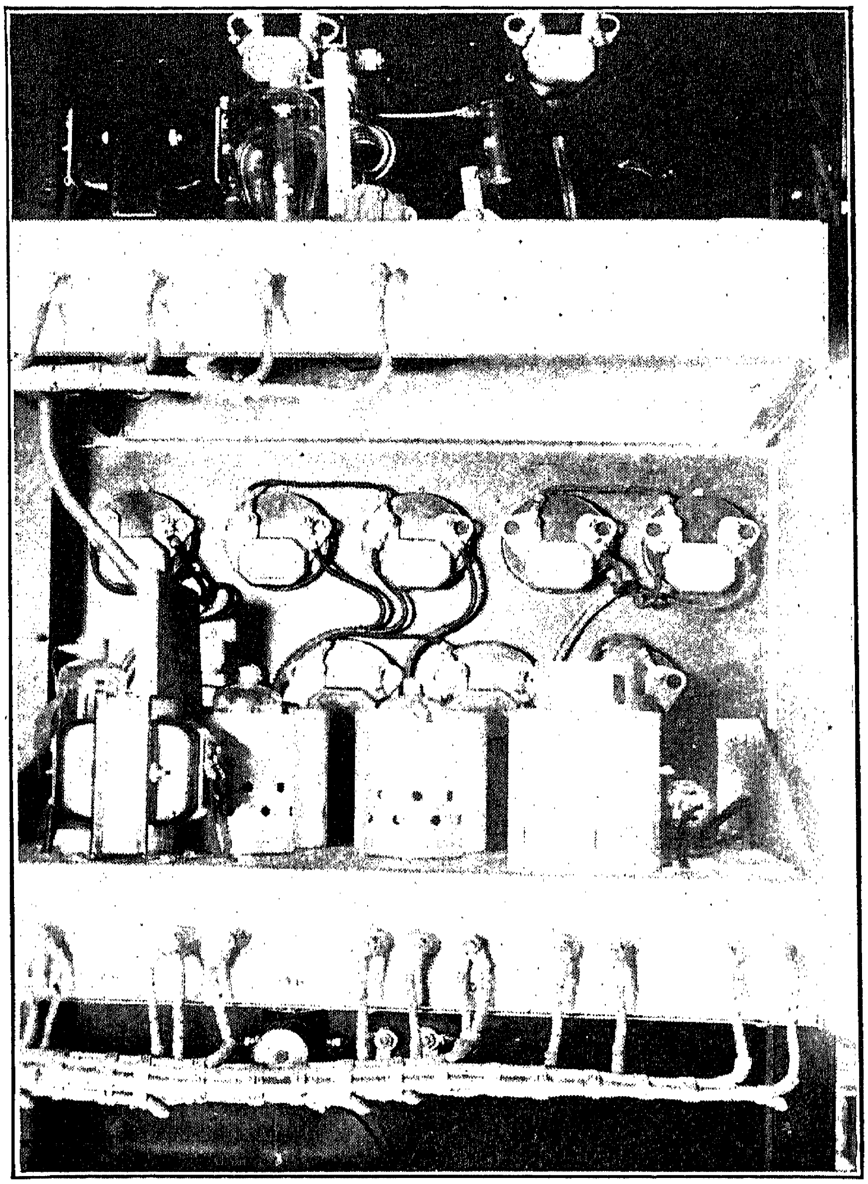


Fig. 19.4.—Harmonic Generators for Carrier Frequency Multiplication (courtesy of NBC).

A frequency multiplier unit is shown in Fig. 19.4, while an RCA 833 carrier amplifier stage feeding the intermediate power amplifier is seen in Fig. 19.5. The output of the pair of RCA 846 tubes serving as intermediate carrier power amplifier supply two grid-modulated RCA 899's, whose plate tank circuit is coupled to the antenna feeder. An RCA 899 amplifier is illustrated in Fig. 19.6, showing the grid-neutralizing sleeve enclosing

the end of the tube and an air blower to cool the metal-to-glass seals. The midpoint of the circuit between the two grids of the final r-f power tubes is connected to the plates of the output tubes of the video amplifier, which are two RCA 891's in parallel, through a bucking voltage supply which is also the coupling impedance. The voltage from the bucking unit is equal to the plate voltage of the video tubes plus the magnitude of the negative grid voltage of the RCA 899's. This voltage supply is regulated, so that the low frequencies are transmitted directly through it, while the upper video frequencies pass through coupling filter circuits. The filter

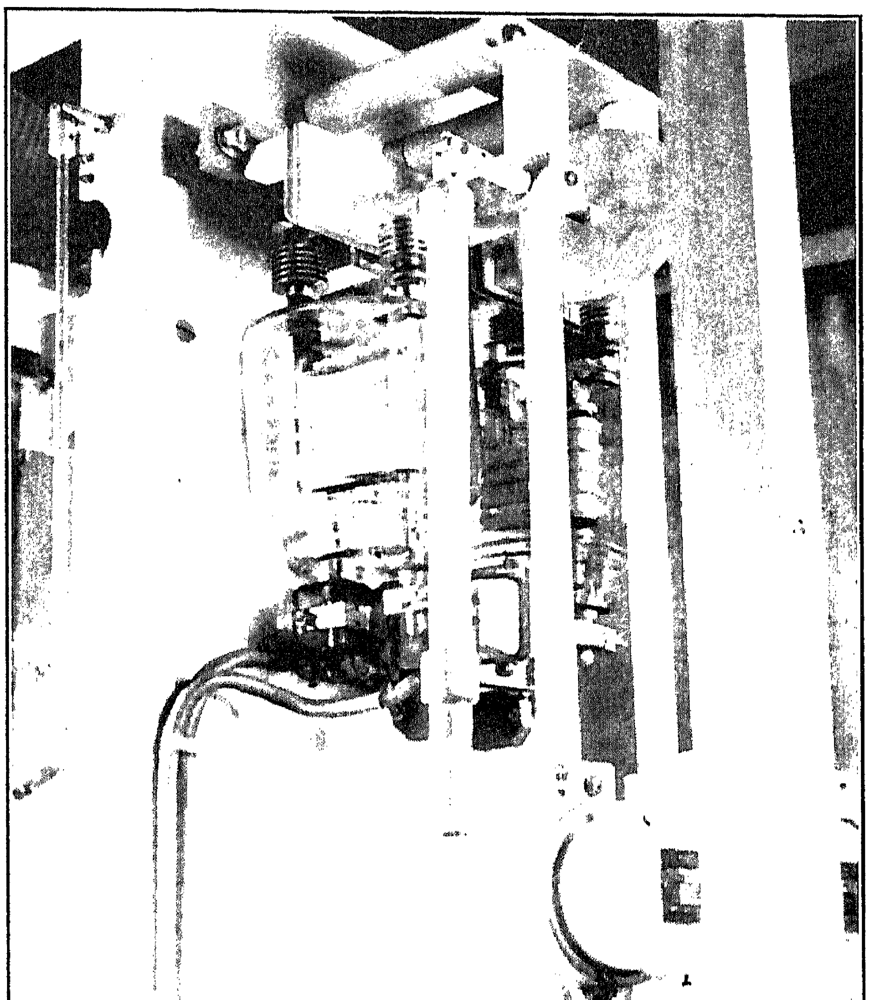


Fig. 19.5.—An RCA 833 Carrier Amplifier Stage (courtesy of NBC).

circuit is actually a constant-resistance network such as is illustrated in Fig. 19.7. The reason for using this method of coupling rather than an ordinary condenser and resistance is that the d-c component of the video signal must be supplied to the r-f tube grids.

The amplifier preceding the modulator consists of four stages. The first stage of the video amplifier contains two RCA 807 tubes in parallel, followed by a stage of three parallel RCA 807's. These two stages are coupled with π networks to separate the plate and grid capacities of the

successive tubes, and the signal traverses them as an a-c signal, the average value of picture illumination being transmitted as pedestal height.

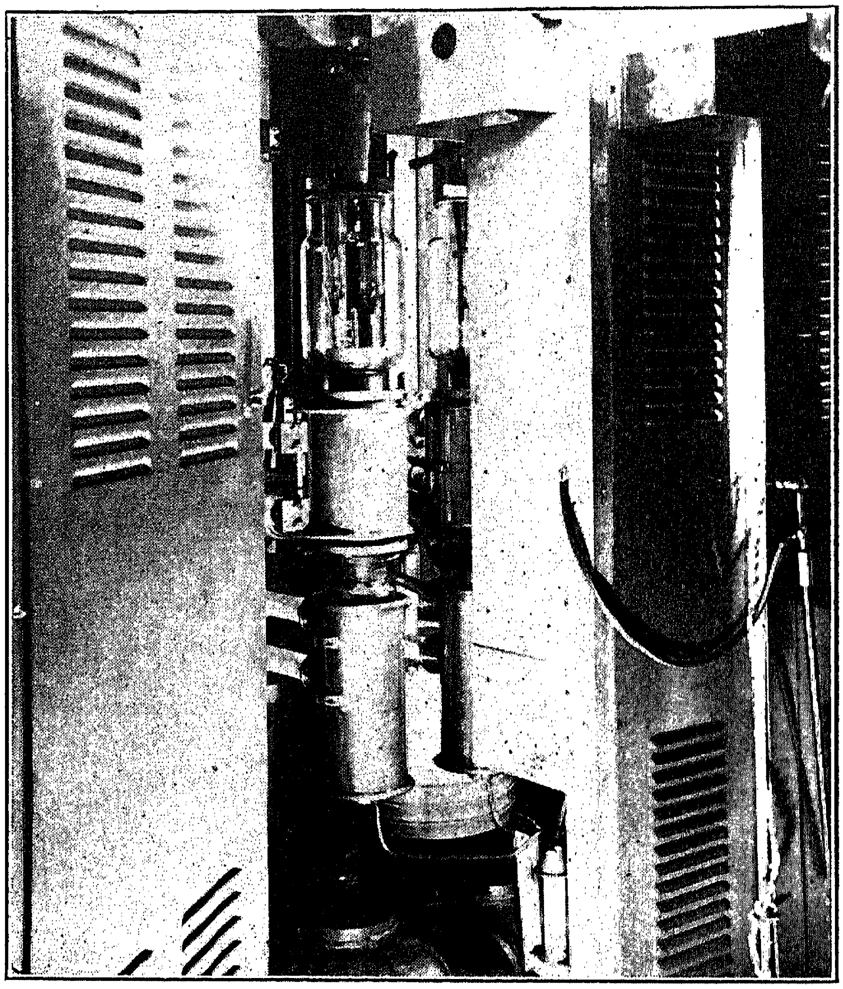


Fig. 19.6.—RCA 899 Output Stage of Empire State Transmitter (courtesy of NBC).

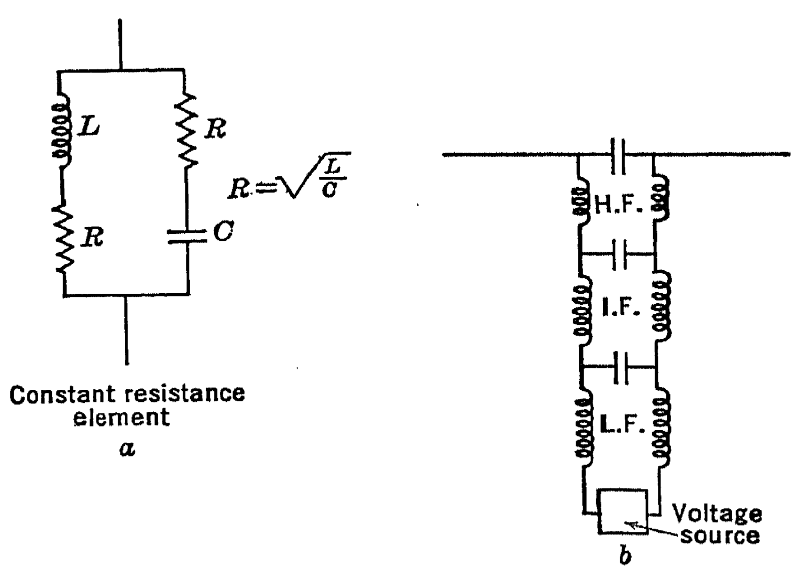


Fig. 19.7.—Constant-Resistance Networks.

The next stage consists of three RCA 831's in parallel. At this stage the d-c component is reinserted by means of an RCA 1-V diode whose cathode

is connected to the grid of the video tube, and whose anode is essentially at ground potential. This stage is coupled to an RCA 891 where the direct current is again reinserted, but since the polarity is reversed the anode of the 1-V is connected to grid and the cathode to ground through the bias supply. The modulator stage consists of two RCA 891 tubes working in parallel as a class A amplifier. The d-c component is, of course, reinserted in this stage with another reversal of polarity. The last three stages use a constant-resistance network as coupling rather than the simple capacity-resistance coupling, because of the necessity of maintaining a constant plate load impedance over the very wide frequency band involved.

The video amplifier of the modulator has to meet the requirement of a flat amplitude response and equal time delay for all frequencies, for reasons that were explained in Chapter 14. The gain per stage in this amplifier is made fairly high in order to economize in tube power. Therefore, interstage coupling is in the form of a four-terminal net which permits the use of a fairly high coupling impedance. This type of coupling has been discussed in connection with the general video amplifier.

Every stage of the carrier amplifier must be adequately neutralized. The push-pull form of this amplifier lends itself to a balanced bridge neutralizing arrangement. Except for the final stage, neutralization presents no particular problem because only a single frequency is being amplified. The neutralizing of the final stage is more difficult because of the wide frequency band that is to be handled. A neutralizing bridge, which requires neutralizing sleeves fitting over the grid end of the tubes, is employed. Furthermore, series condensers compensate for residual lead inductance. Even with this arrangement the neutralization is not perfect because of the physical length of the grid itself, and the nature of the compensation for the short length of grid lead between the effective neutralizing condenser and the grid. This leaves a slight difference in the gain of the stage for the two sidebands representing high video frequencies. The effect is small and relatively unimportant.

The power for the tubes of the two transmitter units is supplied from several rectifiers. For the most part these rectifiers use mercury tubes working from three-phase transformers. The final radio-frequency tubes of the video transmitter are supplied with approximately 9000 volts from a rectifier using six RCA 869A's. Three similar rectifier tubes supply 6000 volts to the final stage of the modulator video amplifier. Other, smaller rectifiers supply the remaining necessary voltages. The r-f output stage and the final modulator stage of the audio amplifier are fed from a 7.5-kilovolt rectifier consisting of six RCA 869A's working from a three-

phase 63-kva transformer. A view of the complete transmitter is given in Fig. 19.8.

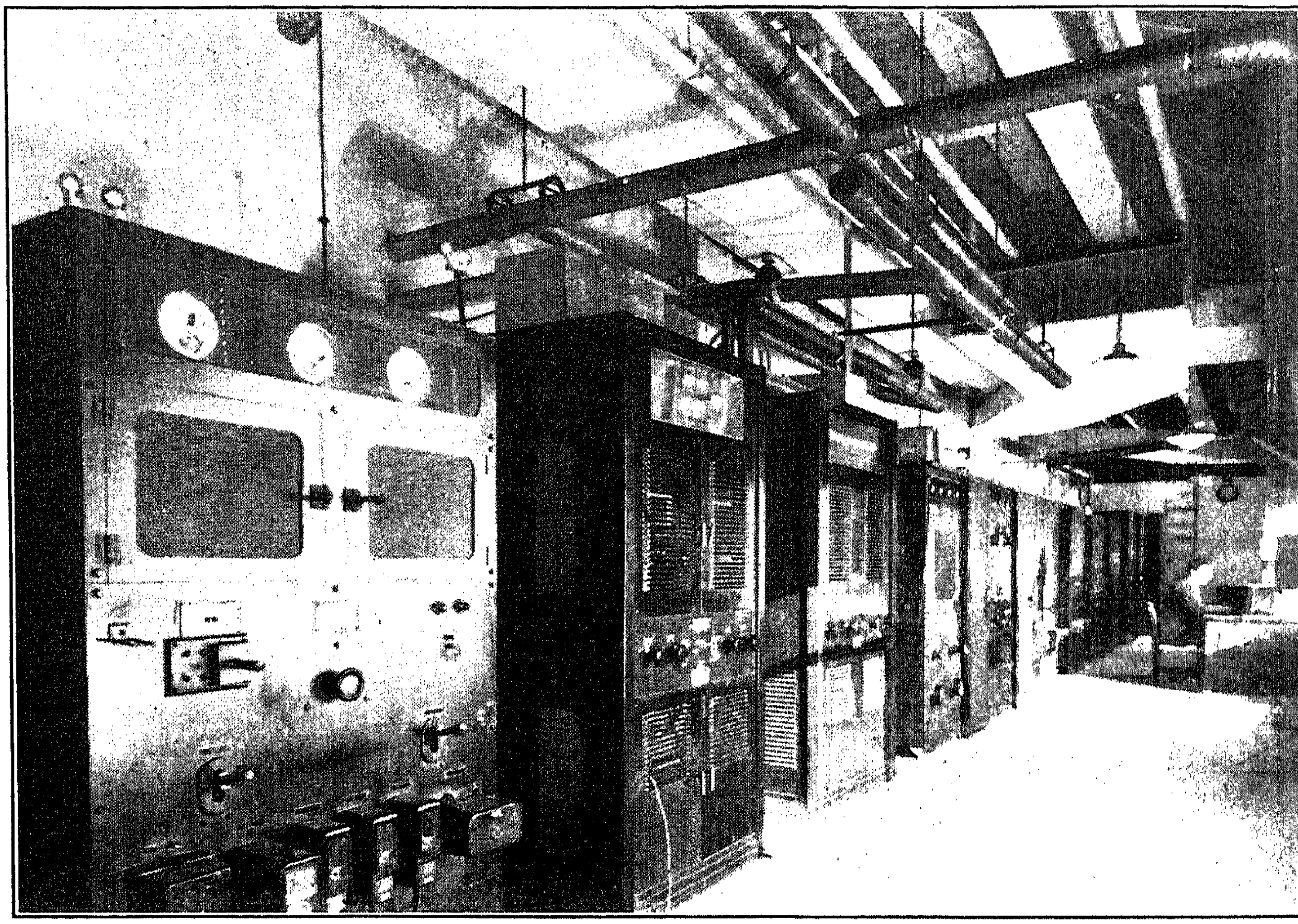


FIG. 19.8.—View of Complete Television Transmitter (courtesy of NBC).

19.4. Special Problems. A great many special problems were encountered in the design and installation of the transmitter at the Empire State Building. Some of these are both instructive and interesting and can profitably be considered in brief at this point.

Symmetric push-pull stages are practically essential where ultra-high frequency is being handled, because otherwise it is virtually impossible to establish a common ground, since the mounting frames, metallic enclosures, etc., are an appreciable fraction of a wavelength in size. Not only was the problem of fixing a ground facilitated by the use of push-pull circuits, but also, by making these circuits symmetrical with respect to their metallic enclosures, the unavoidable stray capacities could be considered as symmetrical circuit elements, thus simplifying the otherwise almost impossible problem of neutralization. At these very short wavelengths, transmission lines become practical as circuit elements, and replace lumped inductances and capacitances as tank circuits, impedance transformers, etc. To illustrate the point in question,

the filament leads on the carrier power amplifier tubes, because of their length, make it impossible to ground the cathodes directly as is necessary in a push-pull stage. However, if the filament leads of each tube are extended so that they are a half wavelength long and are grounded at the far end, the filaments themselves are effectively grounded. In the final power output stage which carries the modulated radio frequency, a similar filament grounding system is employed, modified to accommodate the bandwidth represented by the carrier and sidebands.

As was pointed out in Chapter 16, the effect of the magnetic field of a high-frequency current flowing in a conductor is to cause the current to be forced radially outward so that the entire current flow at the frequencies involved in television transmission is concentrated within a few thousandths of an inch of the surface of the conductor. This property makes it possible to construct concentric circuit elements coated with a thin highly conducting layer, for example, plated copper or silver, on an underlying metal which has the desired mechanical properties, regardless of the specific resistance of the latter. Concentric tank circuits where high-frequency stability is required were made of invar steel having a very low coefficient of expansion and coated with thin silver plate which gives it the same conductivity as though the entire structure were made of silver.

The problem of obtaining suitable condensers for use at ultra-high frequencies was rather serious. The small variable condensers needed for neutralizing and tuning can conveniently be made in the form of a pair of copper disks whose separation can be varied to adjust the capacity. These disks can usually be mounted directly on the circuit elements between which the condenser is to be connected, thus avoiding the necessity of insulating supports. A condenser of this type for the fine tuning adjustment of a 50-megacycle tank circuit is illustrated in Fig. 19.9. The bypass condensers, which usually do not have to carry a high r-f current, can often be standard condensers. However, in some instances the lead inductance of such condensers may be great enough to cause trouble, as for example, by permitting parasitic oscillations. (The use of large parallel disk condensers with very short leads may be necessary at such points.) Such additional condensers were not required in the installation under discussion and generally can be eliminated by proper circuit design. The most difficult types of condensers to find are those used for interstage coupling capacities. The current in such circuits may reach values as great as 30 or 40 amperes, and, because of the relatively low reactive impedance of the tubes themselves, the reactance of the condensers should not exceed 15 to 20 ohms. This means that a capacity of at least 200 micro-microfarads is required which is capable of withstand-

ing more than the applied plate voltage of 10 kv. Condensers with mica insulation might be used, but the dielectric losses were found to be very high at these frequencies, so that a condenser of this type must be operated at a mere fraction of its normal current rating. Air as a dielectric is satisfactory, but air-insulated condensers of this capacity are bulky and do not lend themselves to good circuit design. Compressed-air condensers are smaller and offer promising possibilities. At present, sulphur-insulated condensers have been found most satisfactory. Another promis-

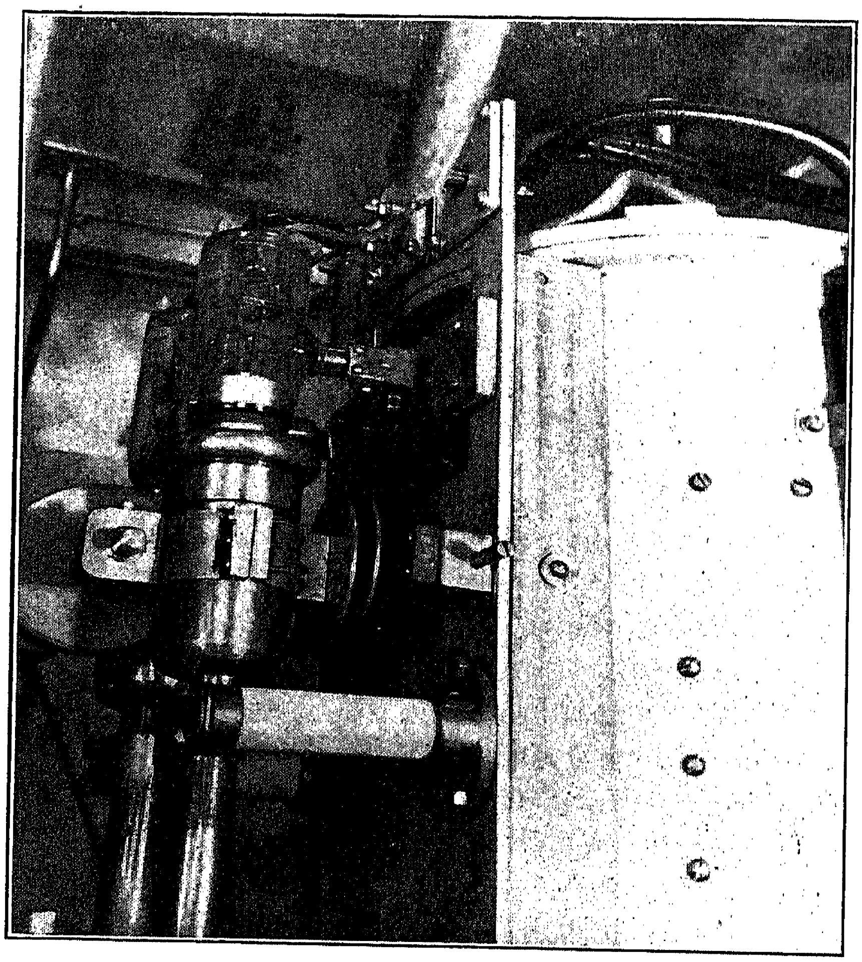


Fig. 19.9.—Disk-Type Variable Condensers.

ing type of capacitor which may eventually prove to be the solution to this problem is the vacuum condenser. The plates of this form of capacitor usually take the form of close spaced concentric cylinders. They are assembled in a glass envelope which is highly evacuated. Thus a high voltage rating can be obtained without the need of large dimensions.

Satisfactory ultra-high-frequency resistors present another serious problem. It is very difficult to make a resistor which does not have a large reactive component at ultra-high frequency. Carbon resistors cannot be used because their resistance depends upon contact between conducting grains, and at extremely high frequencies the capacity of the

grains practically amounts to a short circuit. Thin coatings of a high-resistance metal on glass, or ceramic, have been found suitable for low power. Resistors made by depositing carbon on glass or ceramic cylinders through which water can be circulated, when used as the center conductor of a section of concentric line, have proved satisfactory over quite a broad frequency band and will handle large amounts of power. Also, although of not much importance, a semi-infinite transmission line will have the properties of a pure resistance, without frequency dependence. Finite lines with suitable termination can be made to act as resistances over a fairly wide frequency band.

In the design of the ultra-high-frequency circuits of the transmitter an attempt was made to dispense with supporting insulators wherever possible. However, some insulators are unavoidable. Where these insulators cannot be located at voltage nodes, they present certain difficulties because of the high displacement current that necessarily flows through the dielectric material. Special precautions in the form of avoiding all imbedded points such as screws were required, as the current density in the vicinity of sharp intrusions will be high and the heat developed may cause the insulator to shatter. Electrostatic shielding was also found useful in reducing the current in the dielectric.

The special problems discussed above, together with others which space does not permit describing, have been solved at least to the extent that they do not seriously impair the efficiency of the transmitter in its present form. The limitation imposed by the amplifier tubes, however, still remains. The nature of this difficulty was discussed in Chapter 16. Because of the relatively high capacity of the output tubes, and their low conductance, the efficiency of this stage is unavoidably low. This makes it necessary to use two RCA 899 tubes in the final stages having a power rating of 30 kilowatts each, to obtain the required power output of 7.5 to 10 kilowatts delivered into the antenna feeder.

19.5. Antenna and Transmission Line. The video transmitter is designed to generate two sidebands, each 4.25 megacycles wide. The lower of these two sidebands is partially suppressed by a filter in the antenna coaxial feeder, so that the radiated signal has the frequency distribution shown in Fig. 19.1*b*.

The filter used for sideband elimination consists of six elements. Three of these are of the type described in the text of Chapter 16. These three high-pass dissipative elements are cascaded, as shown in the block diagram of Fig. 19.10*a*. Transmission characteristics of this triplet are shown by the dotted curve in Fig. 19.10*c*. The cutoff is not sharp enough to meet the requirements of the band shape. Therefore three narrow

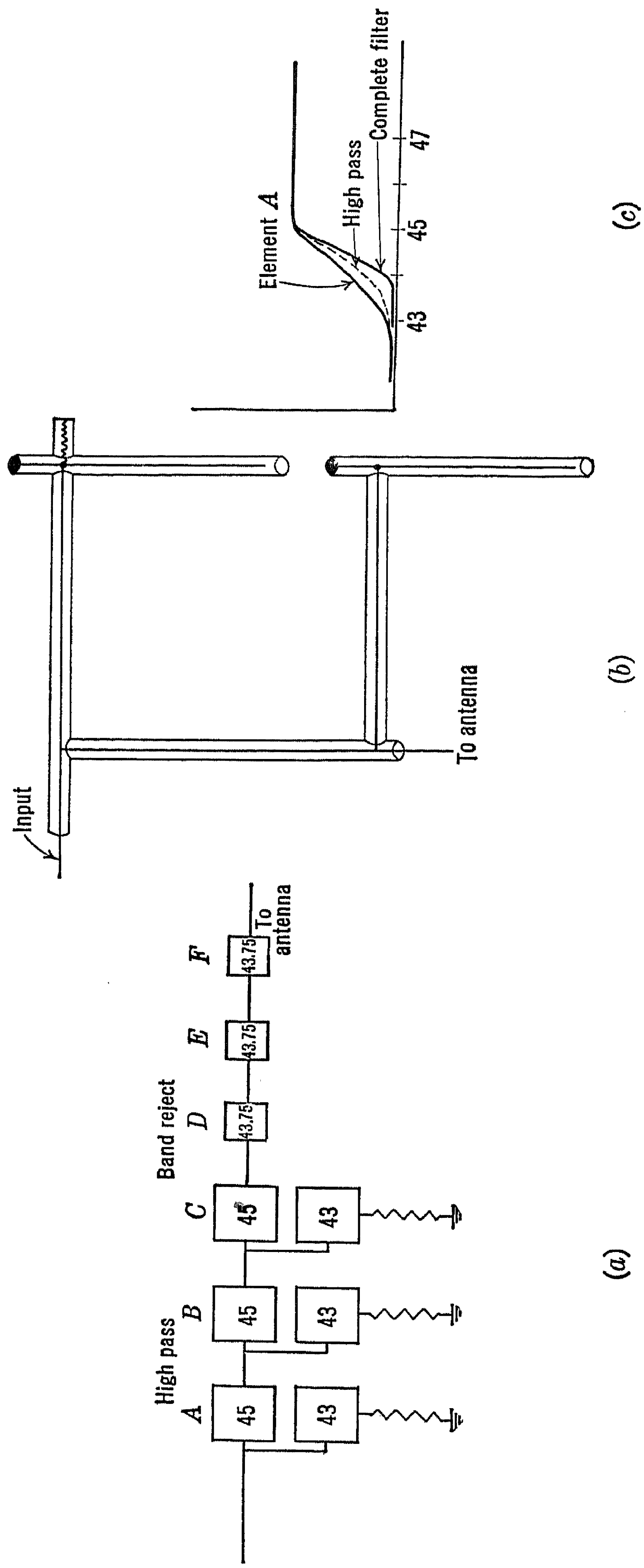


FIG. 19.10.—Coaxial Sideband Elimination Filter.

bandpass elements are needed to complete the filter. The construction of these elements is shown in Fig. 19.10*b*. The overall effect of this filter is included in Fig. 19.10*c*.

A single coaxial line carries the r-f signal to the filter. This feeder has a surge impedance of 72 ohms and an outside diameter of 2½ inches. At the filter output the feeder is divided to form a balanced line, the division being effected by the circuit element illustrated in Fig. 19.11. The antenna itself is formed of two pairs of radiating elements whose axes are

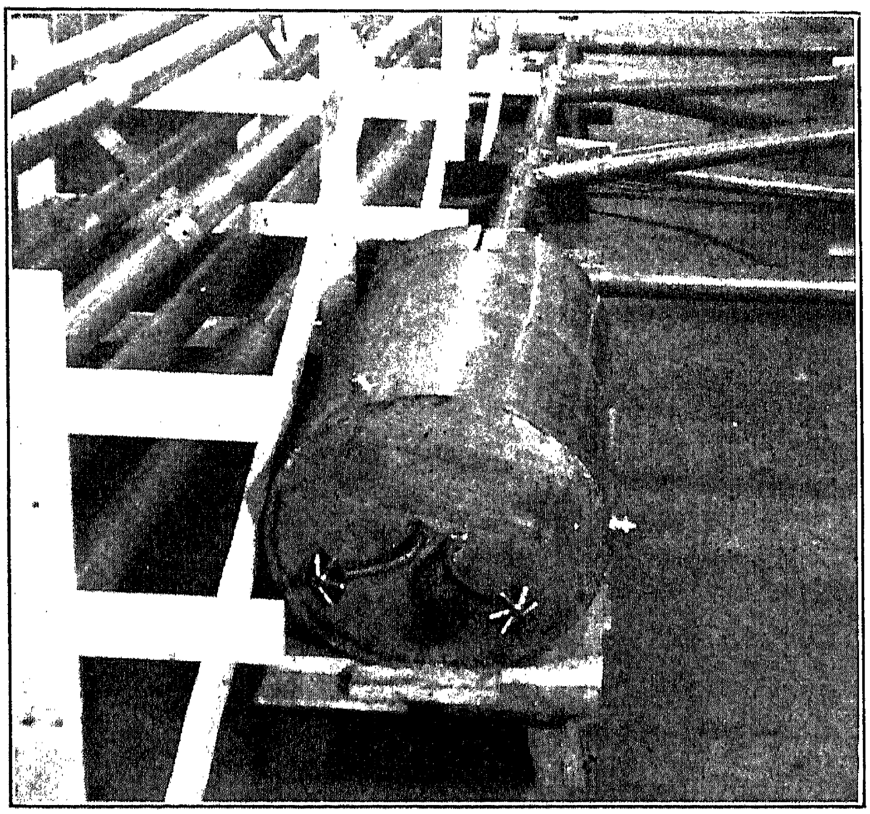


FIG. 19.11.—Photograph of Transmission Line Divider (courtesy of NBC).

at right angles in a horizontal plane. These elements are fed in quadrature through an interconnecting quarter-wave line. The form of these radiators was shown in Fig. 16.37, the complete antenna being illustrated in Fig. 19.12. The inner elements have a maximum diameter of 16.8 inches, and the overall length is approximately 43 inches.

19.6. Video Interconnecting Links, Cable and Radio. It was mentioned in Chapter 18 that two video communication links were provided between the studios at Radio City and the Empire State Building.

One of these is a low-loss cable. This cable consists of two coaxial lines and four line pairs. One of the coaxial lines which go to make up the cable is shown in Fig. 19.13. The central conductor is a No. 12 copper wire. A number of insulating spacers, approximately ¾ inch apart, separate the conductor from the outer sheath. The sheath is made up of

wound copper strip, protected by a layer of steel tape and paper wrapping, the whole being coated with an air-tight lead jacket, making the outside diameter about 7/8 inch. The entire space within the lead jacket

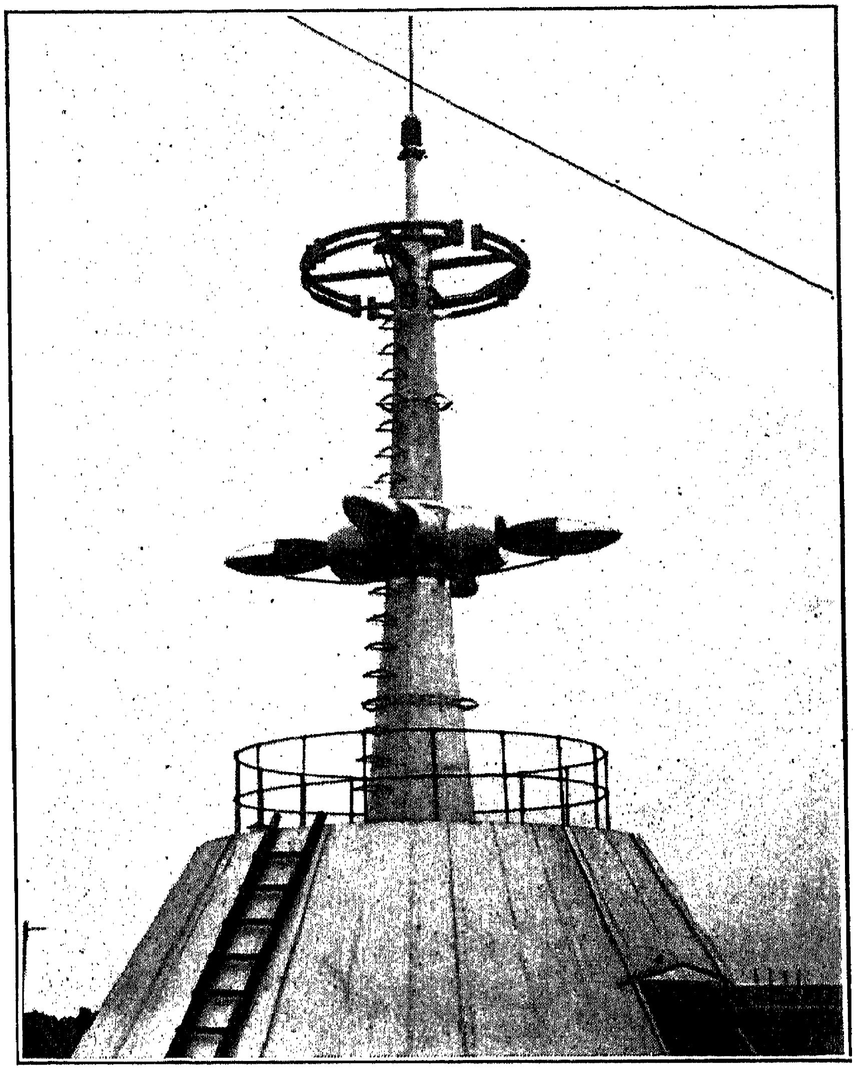


FIG. 19.12.—Special Antenna Used for Television Transmissions (courtesy of NBC).

is filled with nitrogen under pressure, the pressure being indicated by a permanently attached gauge and supervised by an alarm circuit. In this way, possible effects of atmospheric conditions are eliminated.

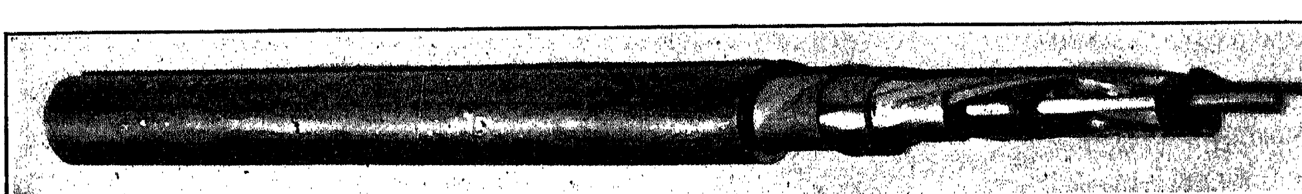


FIG. 19.13.—Low-Loss Video Cable (courtesy of NBC).

An ultra-high-frequency radio relay transmitter and receiver constitutes the other interconnecting means. The carrier frequency chosen for

this radio link is 177 megacycles. The reasons for this choice were the following: First, the frequency must not coincide with possible harmonics of the main television transmitter. Second, it should not exceed that which can be conveniently handled by the available vacuum tubes. Finally, it should be as high as possible, subject to the second condition, so that it can be effectively directed and so that there will be a minimum of interference either from existing radio services or man-made static.

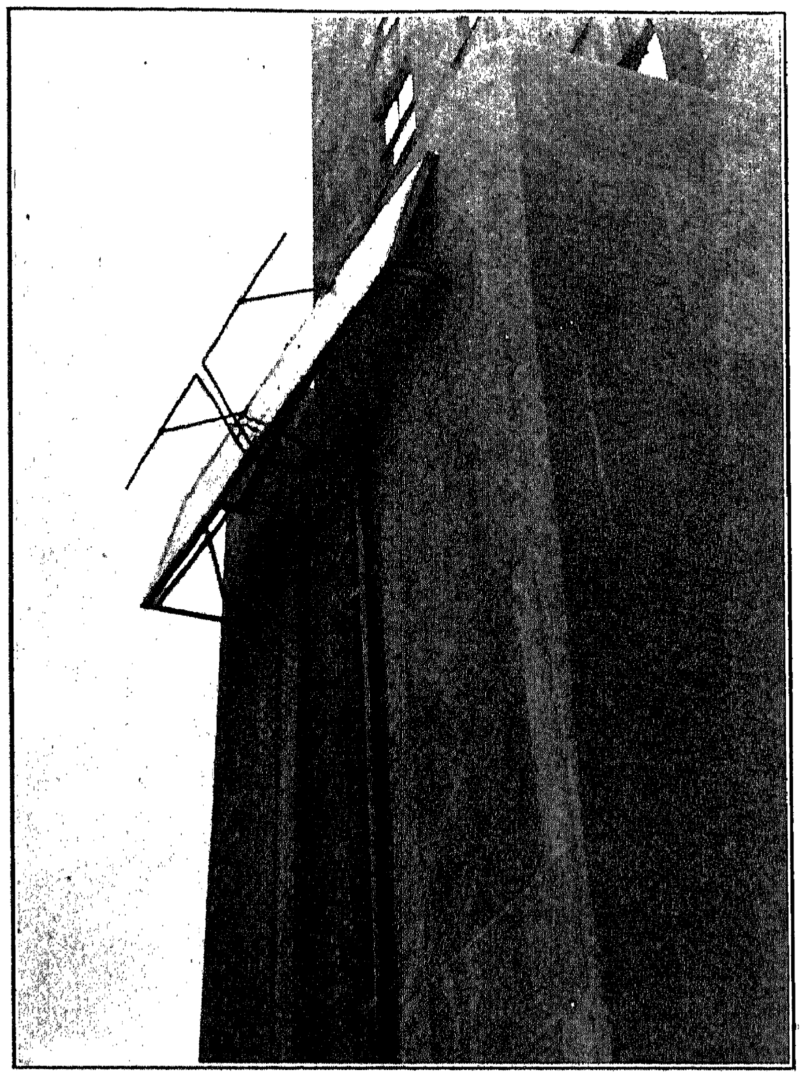


Fig. 19.14.—Transmitting Antenna of Television Relay Link (courtesy of NBC).

The bandwidth of this transmitter was originally designed to accommodate a double sideband signal of 5 megacycles and is now being remodeled to meet the RMA standards.

The link transmitting antenna is located at the level of the fourteenth floor on the south side of Radio City. It consists of two half-wave horizontal radiators fed in phase, mounted on a large copper reflector which gives it the desired directivity. Fig. 19.14 illustrates this antenna. In the north wall of the Empire State Building, at the eighty-fifth floor, so

FIG. 19.15.—Terrain in Vicinity of Television Relay Link (courtesy of NBC)

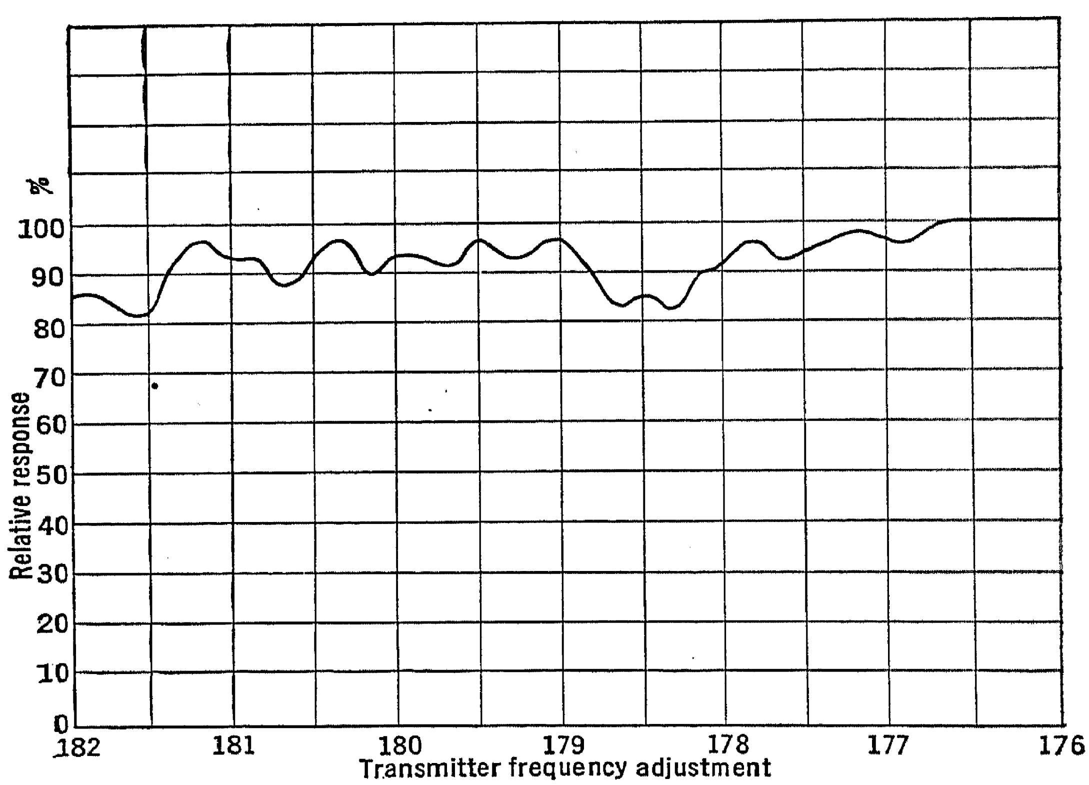


FIG. 19.16.—Variation of Received Signal with Frequency.

placed as to have an unobstructed path to the Radio City end of the link, is the receiving antenna. In construction the receiving antenna is the same as that for the transmitter. The actual distance between these two units is about 4600 feet, or 0.87 mile.

In order to illustrate the terrain over which the relay signals must pass, Fig. 19.15 is presented, being a photograph taken from the position of the receiving antenna looking toward Radio City. In spite of the many possibilities for reflection, the transmitted radiation is beamed suffi-

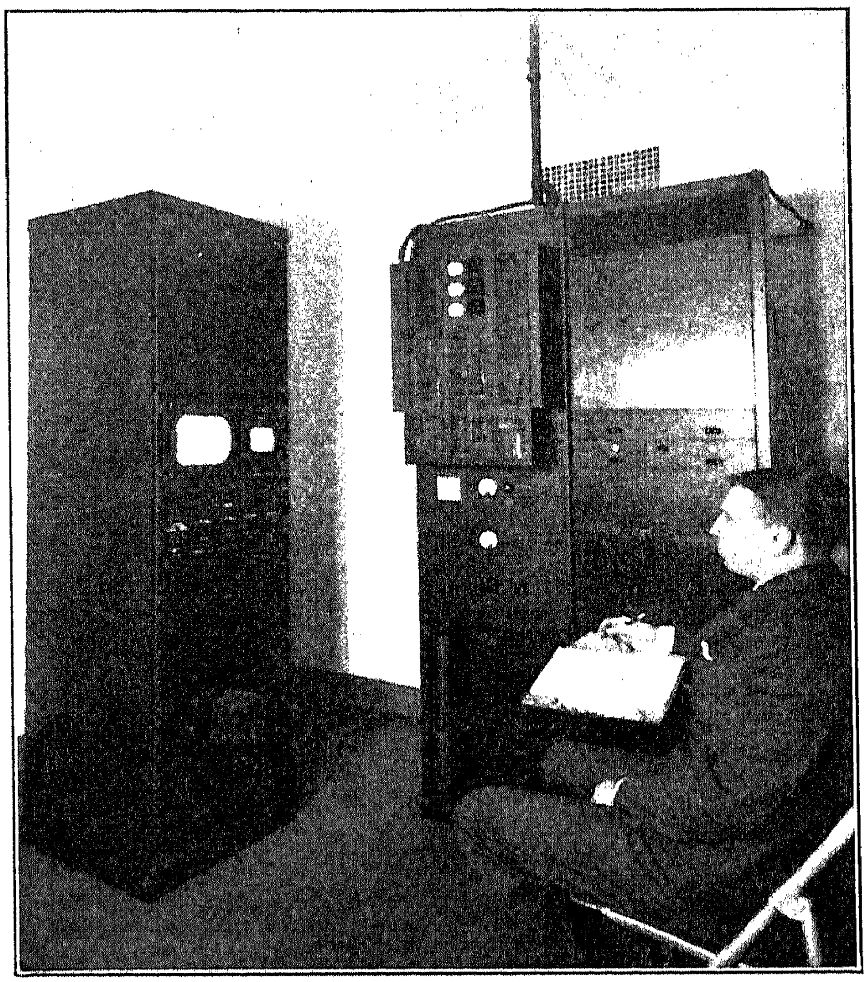


FIG. 19.17.—Ultra-High-Frequency Link Transmitter (courtesy of NBC).

ciently that very little interference of this kind actually exists. The variation of received signal with frequency over a band from 176 to 182 megacycles is shown in Fig. 19.16. It will be seen that it is nearly constant, although a few irregularities do exist owing to the interference effect of reflected radiation.

19.7. Link Transmitter. The transmitter is made up of a master oscillator, video amplifiers and modulator, an r-f power amplifier, and a monitor. A view of the complete link transmitter is given in Fig. 19.17. The panel on the extreme right holds the communication equipment;

next to this, in the center of the photograph, is the rack containing the transmitter itself; on the left is the monitor. The lower and center sections of the transmitter rack house the voltage supplies, and the upper part, the actual transmitter unit.

A schematic circuit diagram of the transmitter is given in Fig. 19.18. The master oscillator is controlled by a concentric tank. The center conducting member is a copper tube 2¼ inches in diameter and 0.2 wavelength long. One end of the conductor is, of course, attached to the outer cylinder by a shorting disk which is silver soldered to both to insure a

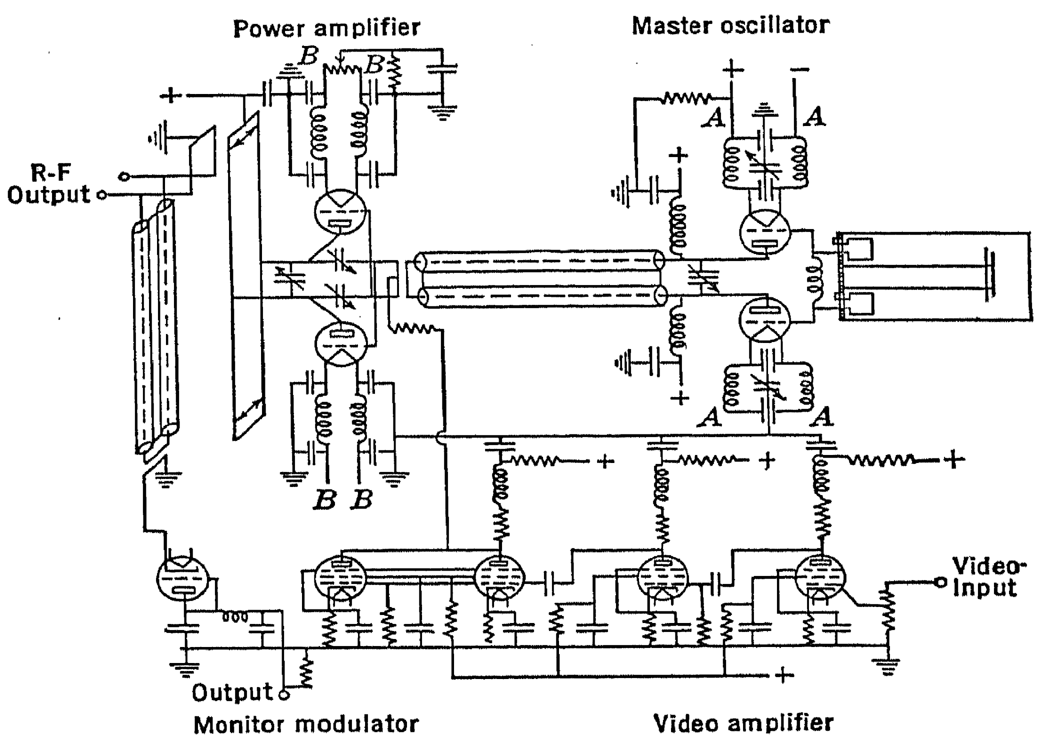


Fig. 19.18.—Circuit Diagram of Link Transmitter.

good electrical path. In order to maintain temperature stability, an invar rod fastened to the disk runs through the center of the inner member. A sylphon bellows about an inch long, which forms part of the inner conductor, is attached to the other end of the invar rod. Since the length of the invar rod is practically invariant with temperature, the effective length of the controlling inner conducting member is constant. The ratio of the diameters of the cylinders is 3.5, and the theoretical Q of the tank is nearly 12,000.

Two RCA 834 tubes in push-pull drive the oscillator, the grids being coupled to the controlling tank by small loops which are inductively coupled to the center member at its current maximum. A short wire whose length can be varied, in parallel with the coupling loops, adjusts the grid reactance to the optimum value.

A concentric line, tuned with a balanced condenser, is connected from plate to plate of the oscillator tubes and forms the plate tank circuit.

The grids of the variable gain power amplifiers, which are also RCA 834's, are inductively coupled to the midpoint of this line. This stage operates as a cross neutralized push-pull amplifier. All leads are as short as possible to avoid unwanted inductances, particularly those of the neutralizing condensers. The stage is so constructed that the physical symmetry is very high. In order to preserve this symmetry, the plate inductance is in the form of two balanced lines in parallel. Coarse tuning adjustments are made by means of sliders on these wire lines, and the fine adjustment by a small movable disk condenser.

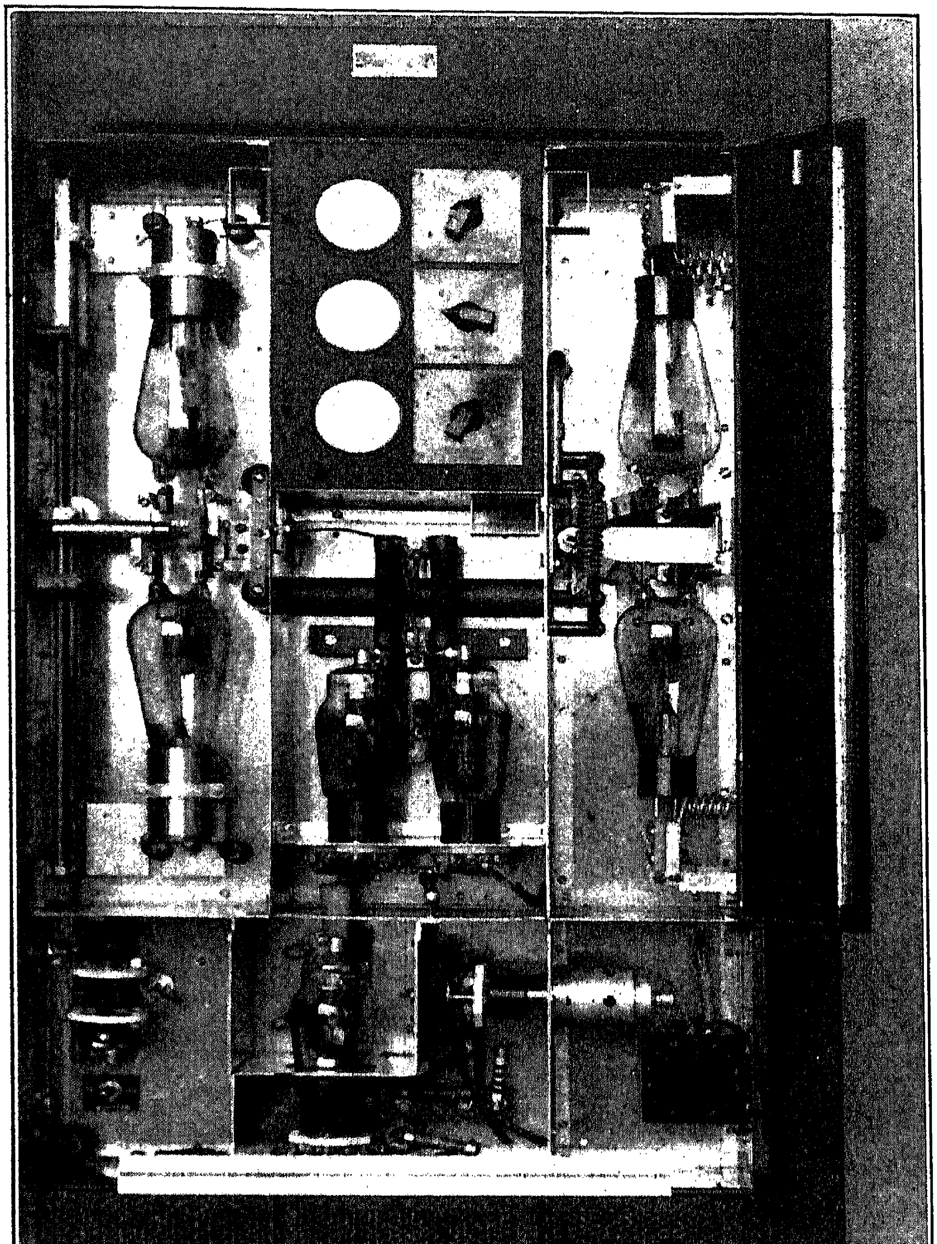


Fig. 19.19.—Interior of Relay Link Transmitter (courtesy of NBC).

ate frequency appearing at the output of this push-pull detector is coupled to a six-stage bandpass amplifier. A diode rectifier in the form of an RCA 955 tube serves as second detector and feeds an RCA 42 output tube. The gain of the intermediate amplifier is controlled by the output of a d-c amplifier which is driven from a voltage divider across the diode load resistance. A manual gain control is also provided. The antenna is located approximately 100 feet from the receivers, and two 76-ohm coaxial lines form a balanced feeder between the two. Like the video cable discussed at the opening of this section, the cable is filled with nitrogen under pressure to make its performance independent of atmospheric conditions. Another concentric line also leads from the receiver to the antenna reflector where it is terminated by a small damped radiating element. At the receiver it is connected to a calibrated oscillator which

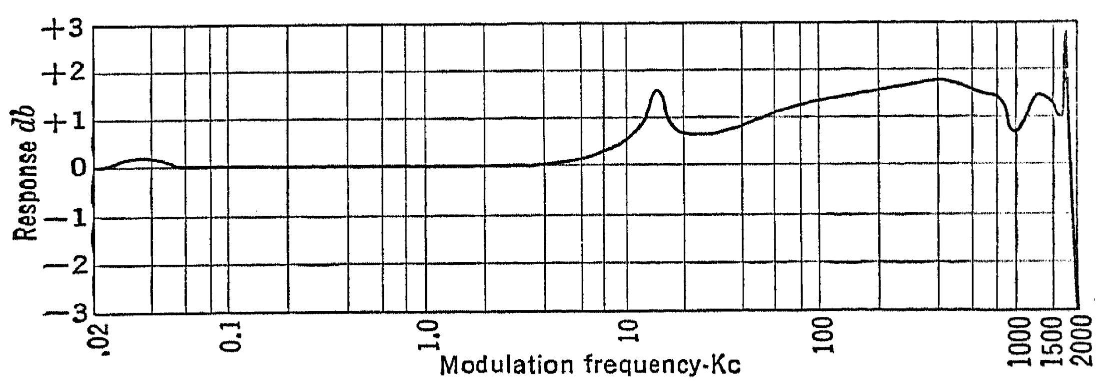


FIG. 19.22.—Overall Response Characteristic of Television Relay Link.

can be varied over the range from 172 to 182 megacycles. The purpose of the arrangement is to enable the operator to have available a signal of known amplitude and frequency with which to line up the various elements of the receiver circuit and to observe the shape of the response characteristic.

Because of the high directivity of the transmitting antenna, a signal of considerable intensity is picked up at the receiver. The field strength at the receiving antenna is calculated to be 30 millivolts per meter. The directivity of the receiving antenna results in permitting little radiant energy arriving from any direction except that towards which the antenna is pointed to affect the receivers. Consequently the signal-to-noise ratio is high. Measurements of this ratio show that the noise level is some 44 decibels below a signal corresponding to 85 per cent modulation of the transmitter. The noise observed was almost entirely due to 60- and 120-cycle ripple from the power supplies. It is interesting to note that even the interference resulting from lightning flashes in the neighborhood of the Empire State Building only produces a moderate click in the receiver output, while the effect on the receiver of such man-

made static as results from the operation of elevators, etc., is entirely negligible. Also no interference is observable at the receiver from the high-power television broadcast transmitter in spite of its proximity.

The response characteristic of the overall relay system is shown in Fig. 19.22. It will be seen to be constant within a few per cent. Pictures relayed over this link are found to be excellent in quality.

19.9. The Mobile Unit. The mobile unit has already been mentioned as an adjunct to the direct pickup and film studios at Radio City. This

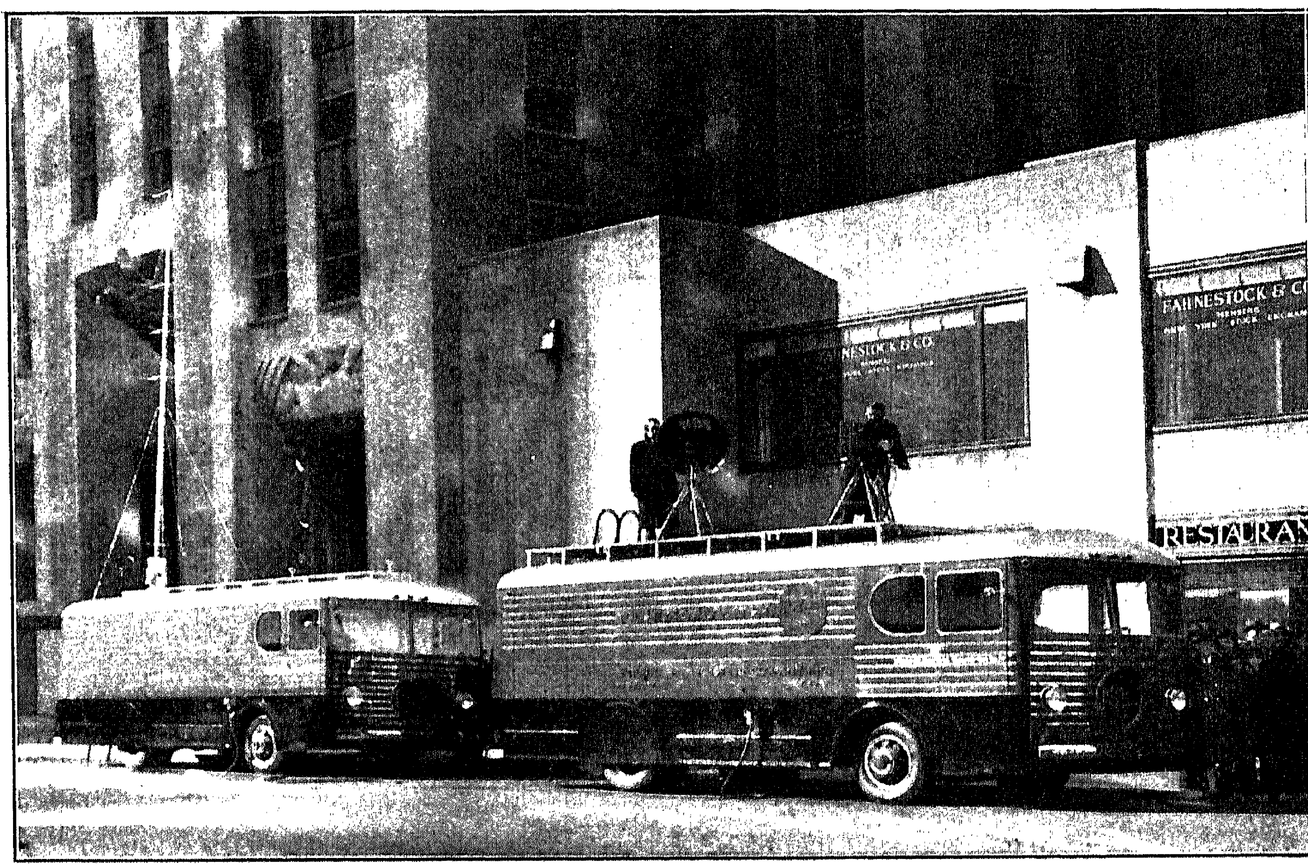


Fig. 19.23.—Mobile Television Unit (courtesy of NBC).

equipment is an essential part of the television project, permitting as it does the direct pickup of events having current interest. The unit consists of a complete video pickup system and an ultra-high-frequency link transmitter.

Two streamline buses, each 26 feet long and capable of a 10-ton load, house the unit. Along the top of the video bus is a catwalk equipped with four camera and microphone placements, which permit mounting the camera and parabolic microphones well above the heads of any spectators watching the scene being televised—otherwise the camera can be placed on the ground at any distance up to 250 feet from the bus. A 24-foot collapsible antenna can be erected on the roof of the transmitter bus. The complete unit ready for action is shown in Fig. 19.23.

A view of the interior of the video bus is given in Fig. 19.24. The first panel on the left contains the audio equipment for the sound accompanying the picture. The next rack contains the monitoring unit. This unit is very similar to those described in connection with the studio equipment. The monitoring rack is followed by one which houses the video amplifier and on which are located the gain, pedestal, and shading controls. Two blank racks follow which will eventually contain a second video channel.

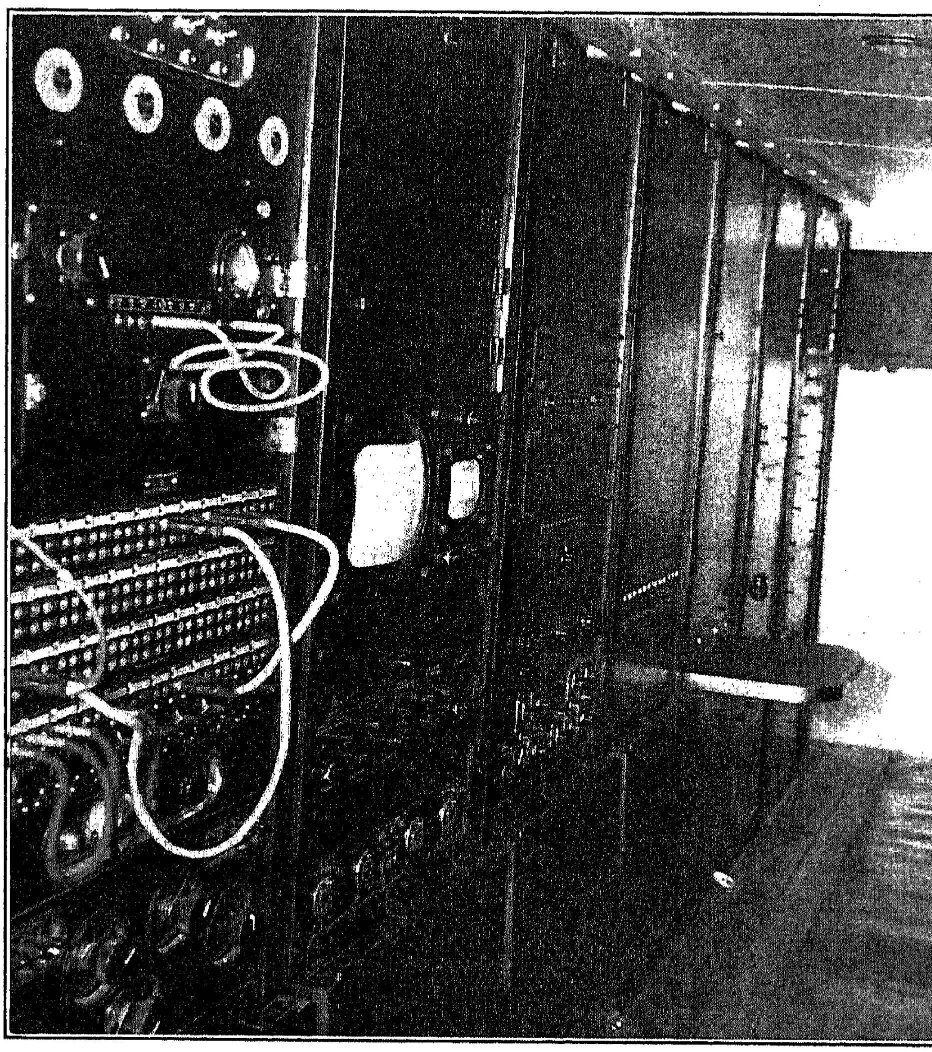


FIG. 19.24.—Interior of Video Section of Mobile Unit (courtesy of NBC).

The final two units are the synchronizing equipment and line amplifiers. The Iconoscope voltage supplies and deflection generators are housed in two more racks on the left not shown in the picture.

The Iconoscope cameras, like the studio cameras, include the video pre-amplifier, a horizontal deflection amplifier, and a vertical blanking amplifier. The camera mounted on its Mitchell tripod is shown in Fig. 19.25. Two hundred and fifty feet of 32-wire cable, which includes a low-loss video line, is available to connect the camera with the bus. Four

microphones, with similar lengths of cable, are provided for the sound pickup.

It is important to have an adequate communication system interconnecting the operators of the various pieces of equipment if a logical, continuous program of the coverage of an event is to be maintained. This is provided in the form of three telephone circuits. One circuit links the video operator with the men handling the camera and other equipment.

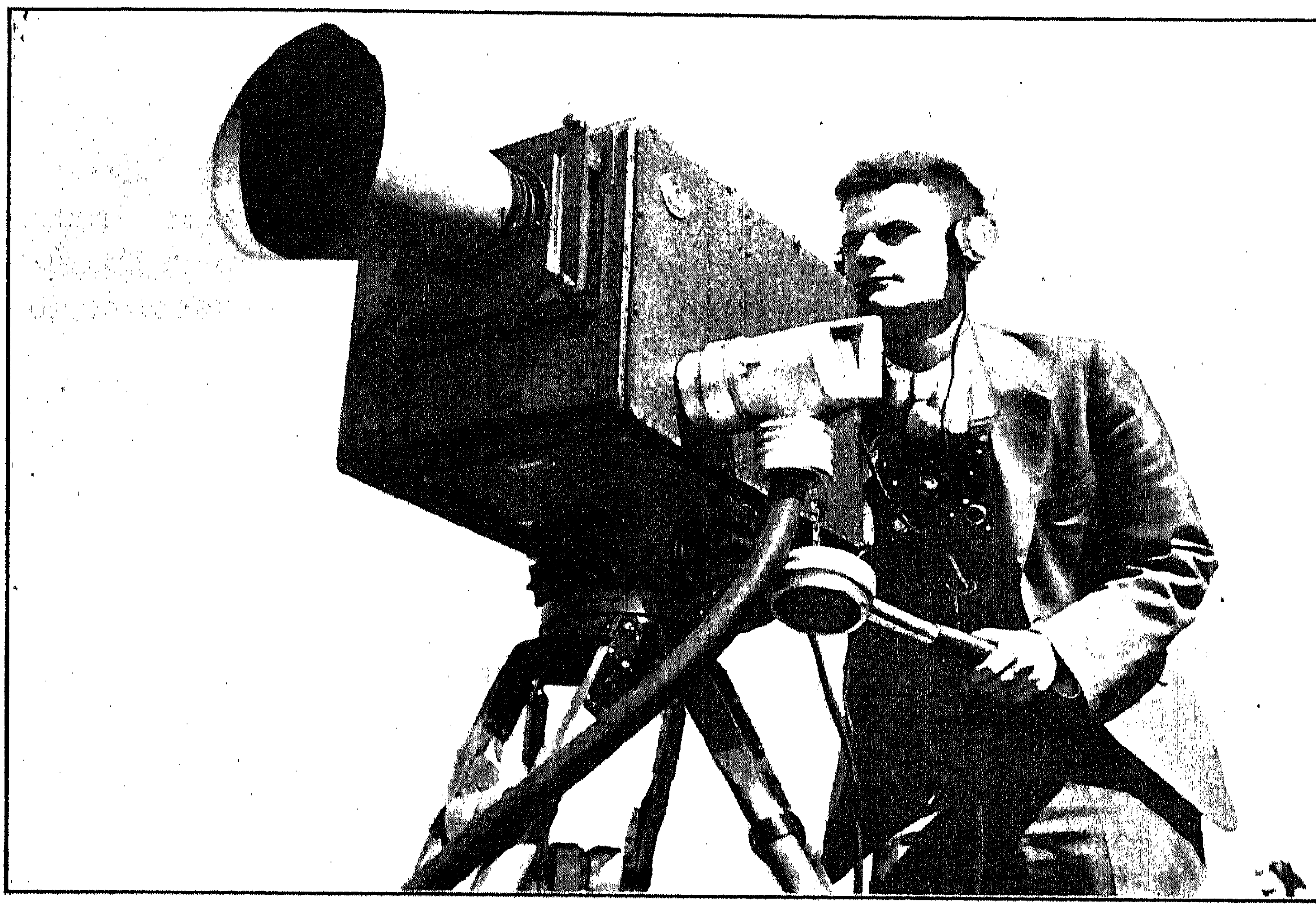


FIG. 19.25.—Portable Iconoscope Camera Used with Mobile Television Unit (courtesy of NBC).

The second connects the audio position with the main studio. These two circuits can be interconnected when necessary. A third line is provided between the program director observing the monitor and the announcers or other staff members at the camera.

The transmitter operates on a frequency of 177 megacycles, using a double-sideband signal corresponding to a video channel of 3 megacycles. A resonant concentric tank controls the master oscillator. The oscillator circuit includes two RCA 887's and like the link unit described in the preceding section develops enough power to operate the power output amplifiers directly. The amplifier stage also uses RCA 887's operating class C,

and is grid modulated by the video signal. A monitor in the output of the transmitter aids in keeping the radiated signal free from defects.

The transmitter is divided into two sections—one containing the r-f equipment and the other the video power amplifier. These are arranged in the transmitter bus so as to be accessible from all sides to facilitate adjustment and repair. The cables for the power needed by the unit and to interconnect the buses are stored in reels at the rear of this bus. They are provided in 60- and 250-foot lengths permitting the sections to be used close together or far separated.

The antenna structure is an example of the high gain and directivity that can be secured at high frequencies in a relatively small space. There are many times when it is not feasible to use the permanent antennas, such as when the bus is surrounded by buildings. Under these circumstances, a dipole and reflector are mounted on the roof of one of the surrounding buildings. An open-wire balanced transmission line is connected from the portable antennas to the transmitter bus.

Tests on this unit have shown it to be very satisfactory. Many interesting retransmissions have been made of scenes picked up from the site of the New York World's Fair and of similar events.

REFERENCES

1. R. R. Beal, "R. C. A. Developments in Television," *Jour. S. M. P. E.*, Vol. 29, pp. 121–143, August, 1937.
2. J. W. Conklin and H. E. Gihring, "Television Transmitter Operating at High Powers and Ultra-High Frequencies," *R. C. A. Rev.*, Vol. 2, pp. 30–44, July, 1937.
3. N. E. Lindenblad, "Television Transmitting Antenna for Empire State Building," *R. C. A. Rev.*, Vol. 3, pp. 387–408, April, 1939.
4. S. W. Seeley and C. N. Kimball, "Transmission Lines as Coupling Element in Television Equipment," *R. C. A. Rev.*, Vol. 3, pp. 418–430, April, 1939.
5. B. Trevor and O. E. Dow, "Television Radio Relay," *R. C. A. Rev.*, Vol. 1, pp. 35–46, October, 1936.
6. P. S. Carter and G. S. Wickizer, "Ultra High Frequency Transmission Between the R. C. A. Building and the Empire State Building," *Proc. I. R. E.*, Vol. 24, pp. 1082–1094, August, 1936.
7 J. Evans, C. H. Vose, and H. P. See, "RCA-NBC Telemobile Units," Presented at the I. R. E. convention, New York City, June 18, 1938.

CHAPTER 20

CONCLUSION

20.1. The Problem of Television Programs. The material used for television programs is fully as important as the technical aspect of transmission. Much time and effort has been spent in determining the type of program best suited for television broadcasts. During the course of the NBC transmission tests a wide variety of programs was produced. These were witnessed on the distributed receivers by audiences consisting of both technical and non-technical observers, and their reaction to the material was recorded. Each observer was requested to fill out a questionnaire asking his or her opinion of the quality, subject matter, arrangement, and general interest of the program. In this way a great deal of information has been obtained in advance of actual commercial broadcasts as to what the public expects and wants from television.

Program material can be drawn from studio productions, current events, and films.

Of the last-mentioned item, little need be said. The high entertainment value, the artistic appeal, and educational qualities that can be attained in a carefully produced motion-picture film are fully appreciated by all. Therefore moving pictures offer very attractive television material, covering as they do a wide variety of subjects including news events, educational features, cartoons, and drama. In addition, films can be used as a valuable supplement to studio productions when the action requires a greater range than can conveniently be obtained in the ordinary stage set.

The problems faced in studio pickup are interesting both in their similarity and in their contrast to the existing vehicles of drama. Like the legitimate theater and sound broadcasting, a television production must be continuous. The intermissions between acts or scenes in a televised play will probably be only short breaks, allowing time for announcements, advertising material, etc. Such interruptions will be far too brief for them to serve as the sole time during which to change scenery. However, since, as in moving-picture practice, two or more stage sets can be employed, this does not offer any serious difficulties.

Television, like moving pictures, lends itself to the use of models. An example of a miniature set for one of the NBC transmission tests is shown in Fig. 20.1. Also lighting effects can be used to their maximum advan-

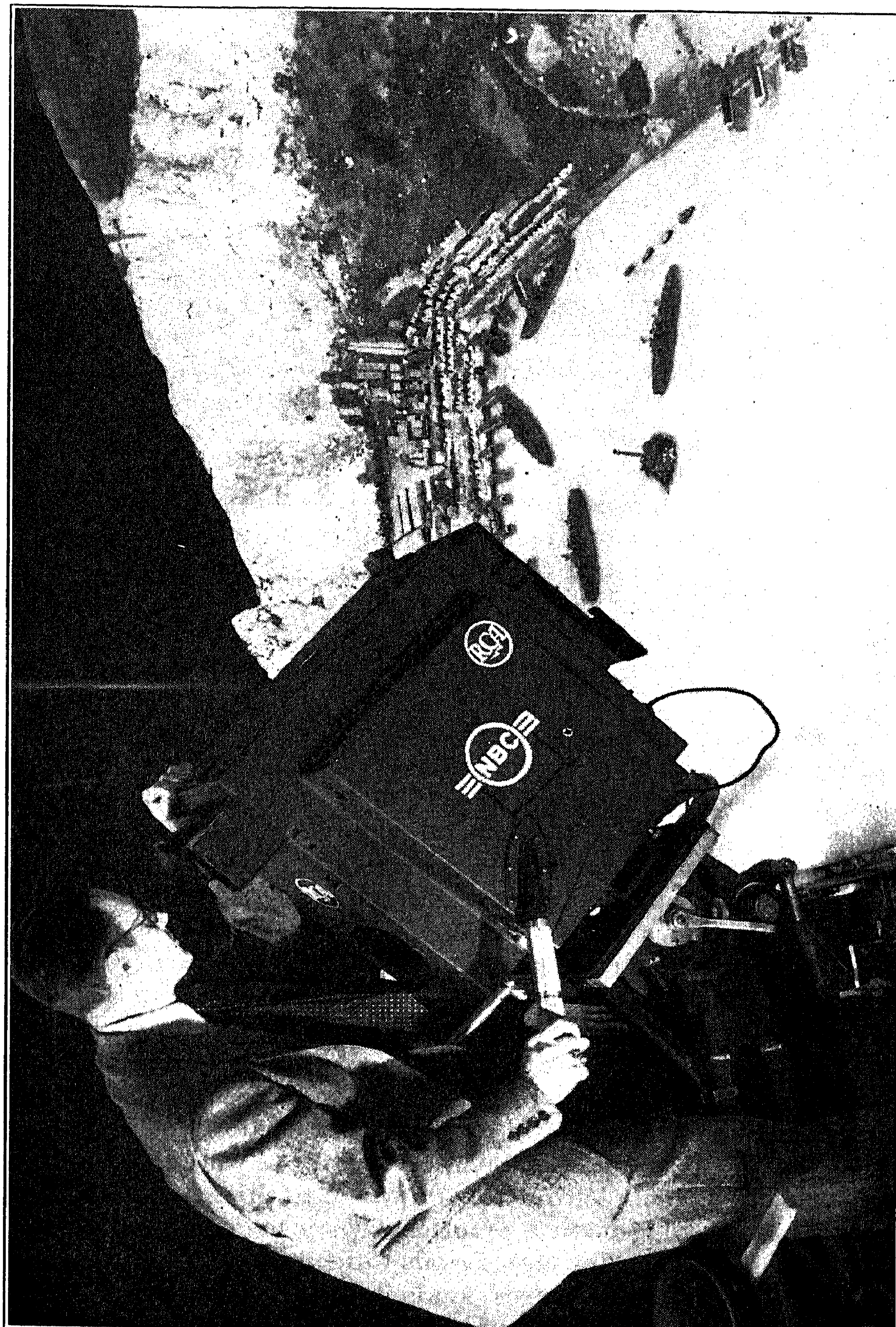


Fig. 20.1.—Miniature Set for Television Transmissions (courtesy of NBC).

tage, because of the fixed, predetermined camera position. Continuity of program places some limitation on the type of stage effects which can be obtained—for example, the instantaneous disappearance of characters or objects encountered so often in movie technique offers some difficulty, although it is not necessarily impossible. The possibility of electrically combining all or parts of two sets, of fading from one set to another, of combining positive and negative pictures, and many similar means of creating illusions make it certain that this new medium of entertainment provides opportunity for many interesting effects when the fantastic is desired. This same flexibility is also an aid in obtaining realism in dramatic production. In this respect the new vehicle much more closely resembles moving pictures than it does the stage.

A television scene must be right at the time of pickup. There can be no "retakes." This means that the ratio of rehearsal time to transmission time must necessarily be great. Experience indicates that this ratio will be in the neighborhood of 20 to 1, and for the more elaborate productions this ratio may rise to more than twice this value. This is to be compared with 7 to 1 for sound broadcasting. However, this is not as formidable as it might seem at first sight because on the average a television program will include a good deal of material which does not require such extensive preparation.

Besides drama itself, studio productions will include musical performances, educational and political addresses, all forms of story telling by one or more persons—to mention only a few of the countless possibilities. All these forms of entertainment can be presented by sound broadcast alone, but their interest will be greatly enhanced if the listeners can actually see the performers—perhaps not watch them continuously, but be able to see when they want to.

The possibilities of television as a means of pickup of current events are almost unlimited. An important difference between films of news events and spot pickup is due to the psychological effect of the time element involved. The film records always make the spectator feel that he is getting the news second hand, whereas instantaneous coverage makes him feel that he is an actual witness to the event. Even the short delay necessary for intermediate film pickup is enough to cause this impression of second-handedness. Direct television coverage of an event does not present this drawback, any more than does watching the event through field glasses or through a window. Thus, in effect, television widens the scope of human experience, just as sound broadcasting has done, by making him more nearly present at the event.

Material for "spot pickup" includes public events, natural phenomena, and occurrences of news importance. Already television has been in a

position to pick up some very interesting happenings—and this without any planned coverage. Sporting events also offer another possibility. Here some fear has been felt that television broadcasts might cause a decrease in attendance. This is probably unfounded; in fact, it is likely to produce the reverse effect by further awakening public interest. Moving-picture recording and sound broadcasting of sporting events have actually increased public attendance. No one truly interested in a sport will be content to watch a televised reproduction if he can attend the event in person. The television pickup is for those only casually interested, and the experience of watching a well-presented reproduction may awaken a genuine interest in the observer.

The problems associated with spot pickup are many, and much remains to be learned about the most suitable technique. Daytime outdoor pickup requires no special lighting—here it is a matter of properly placing camera, microphone, and mobile transmitter; of knowing what will make an interesting picture and what will be merely dull; of being able to select an advantageous site without interfering with other spectators; and many similar items which can be learned only through experience. At night, or for indoor pickup, in addition to the factors mentioned there is the question of lighting. There is no particular difficulty in obtaining sufficient illumination with the portable equipment at present available since the light level required by the Iconoscope is not very high. The difficulty arises in so disposing the lighting units that they will not be objectionable. The few hundred foot-candles incident illumination needed is not in itself objectionable; this much light is almost always present outdoors during the daytime. However, if the illumination in the rest of a room is low, the required light on the subject may produce some discomfort. This is particularly true if the source of illumination is at or below eye level, and the subject has to look directly at it. Time and experience will bring about the necessary operating technique to correct these objectionable details.

The above statements can be made with considerable assurance because they relate solely to the physical functioning of the system. Important as this is, it does not and cannot answer the fundamental question of the psychological and emotional effect of the reproduced picture on the spectator, that is, the question of the entertainment value of a program. This question can be answered only by experience. Sincere effort on the part of the producer, together with cooperation in the form of constructive criticism on the part of the viewing public, will make possible truly desirable programs almost from the start, and will eventually lead to an understanding of the nature of dramatic art as expressed by television.

Television broadcasting will probably go through three phases. The first is the novelty stage, when the picture is watched just because it is a television picture, irrespective of its context. This period will undoubtedly be short-lived and is relatively unimportant. The second phase will be that period during which the broadcast time is limited and the picture is watched more or less continuously by all set owners. This period will be the most difficult from the standpoint of programming. The material must be universal in its appeal, and must be continuously interesting to all spectators. In other words, all programs must have a sustained high level of interest similar to that of the theater. Specialized programs catering to small groups of spectators will be the exception rather than general practice. Meeting this requirement, particularly at the outset, will be very difficult. However, it is by no means an impossible task.

The third and final phase will come when television becomes a more or less general means of broadcasting intelligence. When this phase will arrive cannot be foretold because so many unknown factors are involved. It is unlikely that there will be a decrease in the number of programs having the qualities mentioned in the preceding paragraph, but two new types of programs will be introduced as a result of the more continuous service and the greater number of stations on the air.

The first of these is the specialized program appealing to a more or less restricted group of spectators. These programs may be educational in nature; may relate to a particular sport, hobby, or occupation; or may merely be set apart by virtue of the kind of story they tell. There are unlimited opportunities in this direction, for example, lecture courses, including demonstrations and illustrations; discussions on philately, illustrated by close-ups of the actual stamps; children's stories; lessons in golf, cooking, or theory of television; puzzle contests; and dancing instructions, to mention only a few.

The second new program form is that for which vision is incidental to the sound. Certain musical programs will be among this class, where the audience will not wish to watch the performers continuously, but will want to be able to see them every now and then, just as they would if actually present at the scene of the performance. It will permit the concert listener to see the orchestra leader during the more difficult passages; those listening to popular music to see the singers, solo performers, and the general appearance of the group. In other words, realism will be added to the program. It is not impossible that listeners will also want to see news-casters, sports commentators, etc., and it is certain that the work of the performers will be facilitated by being able to use maps, diagrams, and pictures.

20.2. Television Networks. The limited coverage of a given transmitting station, together with the difficulties involved in chaining stations, due to the line-of-sight propagation of ultra-high frequencies and the great bandwidth of the video signal, will have an important bearing on the development of commercial television itself. Of necessity, it means that at first television will be available only in thickly populated urban centers. Furthermore, the television transmitters in the different centers will be more or less independent of one another.

These restrictions will be reflected in the type of program presented on the broadcasts. The expenses of producing elaborate stage effects will, of course, be greater than if the stations were chained, permitting a sharing of the cost. Therefore there will be a tendency towards simplification in studio production, a trend which will, in the long run, do more good than harm. In the pickup of local events, the very fact that the service area is limited will work to the advantage of the broadcaster. It will mean that material which is not of sufficient national interest to warrant network coverage can be used because of its interest to the local audience.

Films, of course, are little affected by whether the programs are chained or not. In either event they will form a valuable addition to the television program.

The chaining of television stations is a difficult and expensive task, but even with present technical facilities it is not impossible. It is almost a certainty that, as soon as transmission centers have been firmly established in a number of the larger cities, steps will be taken towards forming a network of the stations.

The interconnection between the transmitters may take the form of either coaxial cables or radio relay links. Equally satisfactory results can be obtained with either means, and the choice between them will be on the basis of the cost of installation and operation.

Both methods have been tested experimentally. In 1933 a television radio relay was set up at Arney's Mount, interconnecting Camden, N. J., and New York, to be used in connection with the RCA television field tests. This installation consisted of a superheterodyne receiver, adjusted to a carrier of 44 megacycles. The transmitter operated at a carrier frequency of 80 megacycles and delivered a power of about 100 watts. The modulator amplifier of the transmitter was tuned to a carrier of 8 megacycles and operated directly from the receiver's intermediate-frequency amplifier. A final stage acted as high level detector and modulator. Both receiving and transmitting antennas were highly directional, the transmitting ones having a power gain of 18 in the direction of Camden. This link was used only with a 120-line picture, but the information obtained from it can be applied directly to the problem of a 441-line relay.

During the summer of 1938 the Bell Telephone Company ran a series of tests on the transmission of pictures over a coaxial cable link between New York and Philadelphia. This line was approximately 100 miles in length and had booster amplifiers placed at intervals of about 10 miles. An excellent 250-line picture was obtained with this installation.

The two relays mentioned make use of equipment which is already developed beyond the laboratory stage. It is not impossible that methods which are today in their infancy in the various laboratories, or even as yet unborn, will be ready for practical application when the need comes for laying a television network. Already possibilities are in evidence in the form of radiated centimeter waves and of "wave guides" for extremely high frequencies.

20.3. Television as an Engineering Problem. The groundwork of a practical television system, as outlined in the preceding chapters, has already been laid down. To bring it to this stage of development has required the effort of literally thousands of research men, engineers, and technicians, but this initial effort is small compared to that which will be required once television is established commercially. Those who have followed closely the preceding pages cannot help but be aware of the many unsolved problems in nearly every phase of television. The solution of these known problems will demand the work of a great many trained men. Undoubtedly as television becomes more widespread this demand will become even more acute as a result of many problems which are entirely unrecognized at present.

This aspect of television may well not be the least important of its social consequences. The wide acceptance of television would undoubtedly provide unparalleled opportunities for more trained engineers than can possibly be found at present. How long before the field reaches these proportions cannot be foretold at present, as it is dependent upon many unknown factors.

20.4. Television as a Production Problem. Both the television transmitter and receiver must be considered from the standpoint of commercial production.

The building of television transmitting equipment is not at present, nor will be in the near future, a matter of mass manufacture. Rather each unit is an individual job, being built and tested by skilled technicians, electricians, and engineers.

The receiver, to bring its cost within reason, must be put on a mass-production basis. The organization of efficient assembly lines for a piece of equipment as complicated as a television receiver is no mean task. In the more complicated sets some 3000 parts, 300 to 400 soldering operations, and 25 to 35 vacuum tubes go to make up the chassis alone. Com-

plicated as the construction is, a number of manufacturers have already organized receiver construction on a mass-production basis.

The test routine for television receiving units is one of the most important phases of the production problem. Voltage supplies for the Kinescope or viewing tube and for the amplifier must be tested for voltage, current, and filtering. Both deflection and synchronizing must be checked. Amplifiers, both video and audio, must be aligned and tested. Finally the complete assembled job must be checked under a variety of conditions.

The production of television equipment will also create a demand for engineering talent, trained technicians, and production managers. The task will require the same degree of ingenuity and effort as the developmental problems referred to in the preceding section.

20.5. Television as a Servicing Problem. Servicing is an important part of a coordinated television program. This does not mean that the television receivers are more subject to failure than ordinary radio receivers. In fact, all tests on the experimental receivers placed in the field in connection with the RCA television project indicated that they developed trouble no more frequently than broadcast receivers. The installation and servicing of a television receiver is, however, a much more complicated problem than that of installing and servicing broadcast receivers. Any repair or adjustment which may be necessary, even to such relatively minor items as tube replacements, will probably require the attention of a trained service man. In order to perform this function successfully, it will be necessary to have available a good deal of expensive equipment. Furthermore, this work will require considerable training on the part of the men involved.

Under the circumstances the service organization for television receivers will probably be on a somewhat different basis from the present radio repair work. Coordinated service groups will be formed in the areas where television is available. These organizations will own the necessary test equipment which will be available to their members. The individual service men will thus be relieved of the necessity of a large capital outlay.

Training schools will be required to prepare men for this work. These schools will necessarily be in close contact with the receiver producers and may in some instances be actually operated by the manufacturers themselves.

20.6. The Future of Television. Any statements about what the future holds in store for television and the part that television will play in a future society are obviously pure conjecture. This section is put forward as such. No assumption is made or implied that the authors' contact with the television of today can in any way be of value in predicting

what is in store for this new field of communication. However, this contact, together with the fact that they have seen television grow up from very small beginnings, makes the temptation for them to express their current views on this subject irresistible.

The already published prognostications as to the future fate of television range from suggestions that it will result in a social revolution to predictions of its abandonment in the not too distant future, through lack of interest on the part of the public. Although not impossible, these extremes are the least to be expected. However, merely to say that the most probable future lies along a middle course between these extremes contributes nothing.

Perhaps the best approach is to treat television purely as an improvement in our standard of living in the same sense that sound broadcasting represents an addition to our standard of living. On this basis the general acceptance of television rests on the degree to which it contributes to the standard of living, and the level of the standard of living which our society and economic system can maintain.

Without arguing the point, let it be assumed that our economic system is capable of supporting a widespread television service, if it proves to be of considerable value to the community. This assumption is undoubtedly in accord with fact.

Television will contribute to our standard of living in at least three ways. First, through its cultural and educational value; second, as a source of entertainment; and finally, as an extension of our senses.

The cultural and educational potentialities of television are enormous. Radio broadcasting has, despite any statements to the contrary, contributed a great deal to the cultural standards of our communities. The widespread and increasing interest in forums on current events, in discussions on art and literature, in drama, and in particular in the more advanced forms of music is ample evidence of this fact. It is well known that information can be more easily assimilated through the combined use of the senses of sight and hearing. If, as seems to be true, the broadcasting of sound alone has contributed so much, it follows that when this medium combines sight and sound its contribution will be greatly enhanced.

It has been argued that everything that television can make available in the way of cultural and educational advantages is already available to the public in the form of libraries, lecture courses, art galleries, and museums. This is true, but leaves out of account the fact that the average individual, occupied as he is with the problems of making a living, has only a limited amount of energy which he is willing and able to expend in obtaining access to and absorbing this material. Therefore, if it is

made more readily available through television transmission, he will be able to acquire a much greater knowledge and understanding of this phase of life for the time and effort he can afford to spend.

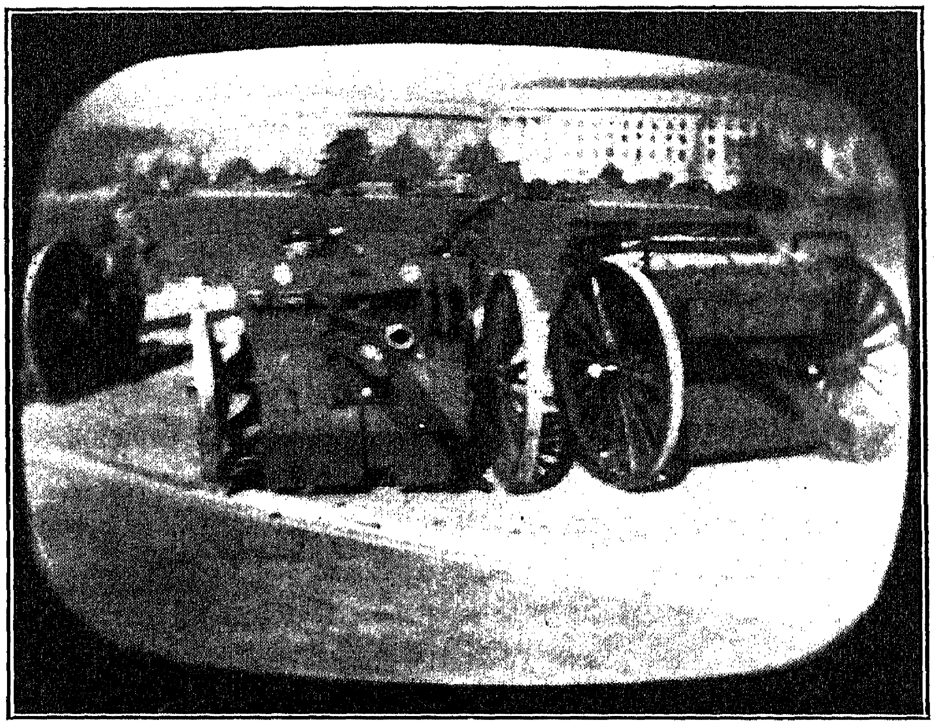


FIG. 20.2.—Outdoor Scene Televised with Mobile Unit (courtesy of NBC).

FIG. 20.3.—Example of Spot Pickup (courtesy of NBC).

The entertainment value of television is evident without further comment. The quality of the pictures reproduced, of course, has a bearing on this aspect. This can best be judged by observing a television receiver

in operation. Three photographs of reproduced studio and outdoor pick-ups are shown in Figs. 20.2, 20.3, and 20.4 to illustrate what may be expected from a home receiver. The apparent quality of the picture, as viewed with an actual receiver, is, of course, greatly enhanced by motion.

Not the least important contribution of television will be its function as an extension to our senses. Broadcasting, as has often been pointed out, extends our sense of hearing immeasurably. It makes it possible to

FIG. 20.4.—Typical Studio Picture (courtesy of NBC).

be present as auditors to most of the important events of our times. Television will permit us to be present as witnesses as well as auditors.

Many programs will be watched from start to finish with close attention. However, there will also be many programs when the user of the receiver will merely listen to the sound and will look at the picture only every now and then when his attention is particularly attracted. Television will enable him to see whenever he so desires. As soon as there is general recognition of the fact that a radio receiver need no longer be blind, the acceptance of television is inevitable.

INDEX